THE JOURNAL OF PARAPSYCHOLOGY

Volume 13 SEPTEMBER, 1949 Number 3

B
OCT 12 1949
PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

Editorial PAGE

Parapsychology and Psychiatry ... 143

Further Work of Dr. Stuart on Interest Test Ratings and ESP ... 151
BETTY M. HUMPHREY

"Direct Influence" of a Person upon the Behavior of Animals ... 166
W. BECHTEREV

A Mechanical Scoring Technique for Testing GESP ... 177
S. DAVID KAHN AND ULRIC NEISSER

Minor Articles

Report on an Experiment Concerning the Influence of Mind Over Matter ... 186
E. A. G. KNOWLES

A Clairvoyance Game Experiment ... 197
WILLIAM R. BIRGE

Letters and Comments

Parapsychology and Marxian Materialism ... 208
ANONYMOUS

Glossary ... 212

A SCIENTIFIC QUARTERLY DEALING WITH EXTRASENSORY PERCEPTION, THE PSYCHOKINETIC EFFECT, AND RELATED TOPICS

DUKE UNIVERSITY PRESS
DURHAM, N. C.

The Journal *of* Parapsychology

EDITORIAL STAFF

J. B. RHINE, J. G. PRATT, and BETTY M. HUMPHREY, *Editors*
J. A. GREENWOOD and T. N. E. GREVILLE, *Statistical Editors*
DOROTHY H. POPE, *Managing Editor*

ADVISORY EDITORIAL BOARD

HORNELL HART
Duke University

H. H. PRICE
University of Oxford

GARDNER MURPHY
College of the City of New York

ROBERT H. THOULESS
University of Cambridge

This JOURNAL is published on the fifteenth day of March, June, September, and December by Duke University Press, Durham, North Carolina.

Contributions submitted for publication and all editorial communications should be addressed to the Managing Editor, Dorothy H. Pope, College Station, Durham, North Carolina. Correspondence with the editors is advised before submitting articles other than reports of experimentation. Since it does not forward manuscripts by registered mail, the JOURNAL cannot guarantee that they will not be lost in transit, and contributors are urged to keep copies of their papers. All contributions should be typewritten, double spaced. References should be given in the form adopted by the JOURNAL.

Reprints may be ordered when the proof is returned. The JOURNAL will bear one half the cost of reprints up to two hundred copies. Correspondence concerning subscriptions, change of address, back numbers, and other business communications should be addressed to Dorothy H. Pope, College Station, Durham, N. C.

The subscription price is $5.00 a year; single current numbers $1.25. The rate for back volumes is $6.00; for single numbers $1.50. Missing numbers will be supplied free when lost in the mails if written notice is given within one month of the date of issue. All remittances should be made payable to the JOURNAL OF PARAPSYCHOLOGY.

Entered as second class matter May 21, 1942, at the post office at Durham, N. C., under the act of March 3, 1879.

DUKE UNIVERSITY PRESS

The Journal *of*
Parapsychology

Volume 13 SEPTEMBER, 1949 Number 3

EDITORIAL

PARAPSYCHOLOGY AND PSYCHIATRY

THROUGHOUT the history of psychiatry there has always been a moderately large number of psychiatrists interested in psi phenomena. As a consequence profitable relations have developed between the two fields with many opportunities for collaboration. A brief review of these relations and opportunities is therefore in order, although the remarks that follow here are not made with any presumption of finality.

Parapsychology is, to a great extent, a historical derivative of psychiatry. If we go back to the time of Mesmer and his followers we find that phenomena such as are now included under extrasensory perception were not uncommonly encountered in the therapeutic practices of the early mesmerists. Mesmer himself, Puységur, Esdaile, and Elliotson, among others, mention these exceptional powers which, in the eighteenth century, were simply taken to be part of the mesmeric effect itself. A hundred years later, in the last quarter of the nineteenth century—a period identified with the work of Janet, Richet, and other French physicians—mesmerism had become hypnotism, and telepathy and clairvoyance had begun to be isolated as phenomena apart from the hypnotic trance and to be left to the societies for psychical research. In the twentieth century, however, some of the psychoanalysts took an incidental interest in telepathy, though now without any association with hypnosis being involved. The boldest of these was Stekel, with Freud and some of his followers taking some interest in the matter and with a deep,

though less explicit, interest on the part of Jung. This attention to the phenomena of telepathy still persists today, especially among the more psychoanalytical psychiatrists, as the names of Ehrenwald, Eisenbud, Servadio, Pederson-Krag, Fodor, and others may be cited to show.

Psychiatric interest in parapsychology was represented in England by such men as T. W. Mitchell, McDougall, William Brown, and Bendit. In this country the development of a medical section of the American Society for Psychical Research, its membership consisting largely of psychiatrists, is noteworthy. We may say, then, that parapsychology has received a fair amount of fostering from its better-established neighbor.

* * * * *

The relation, however, has not been merely a historical accident; there has been reason for the long association of psi studies with psychiatry, even though the shifts we have indicated show that the precise relation has not always been understood.

The phenomena of both fields are, of course, deviations from the normal in the sense of the usual and familiar. Furthermore, many of the occurrences typical of the two are, on the surface, almost identical. Certain types of spontaneous psi experiences, for example, are removed from classification as psychopathological only by what is called their veridicality; that is, their contribution of facts which the subject could not normally have known. The person who has recurrent dreams of the death of a friend may well consider his experience a morbid one unless his friend actually fulfills the requirements of the dream in sufficient detail to rule out coincidence.

Or again, an individual may be regarded with concern by his psychiatrist if he alone repeatedly sees an apparition of a deceased friend; if, however, two or three others see the same apparition and there is a dog or child present who behaves as if seeing the same phantom, the problem shifts from the psychiatric to the parapsychological realm. And so we might continue over a considerable range of cases drawn from the two fields that show similarities on the level of the subject's experience.

These similarities, however, are the more superficial ones. The one clearly conjunctive feature which parapsychology and psychiatry share is their primary preoccupation with unconscious mental life.

This is a definitive characteristic of these two fields and perhaps of these two alone. Psi capacities would doubtless have been described in Aristotle's psychology were not these capacities obscured by their unconsciousness. On the side of psychopathology, as Freud had the genius to appreciate, it is through the relegation of maladjusted desires to the "unconscious" that they are enabled to survive and endanger the happiness of the individual. Accordingly, all the vast psychology of unconsciousness, with its peculiar properties, devices, and functions, is common background for these two branches of the study of man.

We need to ask, however, how basic this association is between the fields we are considering. Is it merely a superficial one based on similarity, or is there some underlying causal relation between ESP and the pathological devices it seems to approximate? Are psi phenomena abnormal or are mental illnesses products of psi effects? A beginning may be made on this important question by reviewing here the approaches that have already been made to the study, but it is emphatically only a beginning. First, the studies of ESP ability by card-test methods conducted in mental hospitals have not thus far established any association between pathology and ESP capacity. Thus far, too, there is no record that paranoid patients who claim they are being persecuted telepathically really possess any special ESP ability. Even from the wide-ranging collection of spontaneous cases there is no evidence of any tie-up of ESP with mental disorders; rather, the impression is given that people who have spontaneous ESP experiences are no less likely than others to be normal healthy individuals.

The general impression from the personality studies of Stuart, Schmeidler, and Humphrey also favor the view, with some reservations, that the better-adjusted individuals are likely to show the largest deviations from chance averages in their ESP data. This relation seems especially clear from Schmeidler's work with the Rorschach test, and there is some unpublished work on card tests with college students which shows a positive correlation of mental adjustment with performance in the ESP tests. Doubtless if there were any basic tie-up of mental illness and psi capacity, the psychiatrist would have encountered more psi phenomena, and parapsychology would likely not have become an independent branch of inquiry.

Recently a new line of work has developed, linking psychiatry and parapsychology and suggesting that the ESP ability of mental patients might be improved by treatment. The rather striking though still unconfirmed experimental results of Urban and Köck at the Neuropsychiatric Clinic at Innsbruck show an increase in ESP scoring rate by patients who have been given narcosynthesis and shock treatment. This study deserves careful attention but one must also keep in mind the possibility that the increase in scoring rate is an effect of the treatment as such rather than of relief from the mental disorder. Urban and Köck have drawn no conclusions.

Altogether there is no good reason to think of psi phenomena as abnormal, nor of mental disorder as favoring psi processes. They have enough in common for many important inter-relationships, but so far none of these appear to be causal. Like any other ability, psi can become involved in pathological thinking—can even become the conspicuous feature. Also, the similarity between psi experiences and pure hallucination often leads to grave errors of judgment that can be seriously unhealthy if allowed to go too far. In such abuses of good judgment, however, the patient is likely to be only utilizing psi as a neurotic device, just as vision or memory or any other normal function can be used in a pathological way.

* * * * *

There are, however, ample grounds for cooperative interest and collaborative research between psychiatrist and parapsychologist, even if no fundamental connection is ever found between psychopathology and psi abilities.

The psychiatrist, because of the nature of his work, has an extremely good opportunity to encounter psi experiences and to make a valuable contribution to parapsychology. Dr. Jan Ehrenwald has pointed this out in the March issue of this year's JOURNAL. Speaking of what the psychiatrist can do with regard to psi phenomena, Dr. Ehrenwald said: "He can study the psychodynamics of so-called spontaneous phenomena; that is, he can apply to them the established principles of the psychoanalytical approach.... Secondly, he can try to arrive at certain generalizations about the modus operandi of the phenomena; thirdly, he can seek to reconcile his findings with the current system of psychiatric thought and to formulate a working hypothesis for further research."

Psychiatrists who make special use of dream material have been finding instances of ESP, mostly cases of telepathy, in the course of the analysis. Some of these cases have already found their way into the literature of psychiatry. Whether the telepathic character of an experience will be noted when it appears depends, of course, on the interest range of the psychiatrist, and the same is true with relation to other forms of ESP. The psychiatrist's opportunity to interpret the patient's psi experience in terms of the knowledge acquired concerning his personality and case history is one that is available to no other student of parapsychology. We should therefore expect the best insights into how ESP functions in the natural life of the individual to be obtained in this way, although because of its clinical nature such research must be exploratory in type rather than conclusively evidential. The exploratory function in science is just as important, however, as the verificative, and just now is more in need of expansion.

We hope, quite naturally, for a wider program of search for ESP effects by psychiatrists. If the kind of work instituted by Urban at Innsbruck continues and broadens, as we hope it will, there will be comparative tests of psi capacities at different stages of all the various therapeutic procedures now in use. It is understandable for Dr. Urban to have begun his inquiry with the more medical methods of therapy, but the same comparative tests could be carried out at regular intervals with the purely psychotherapeutic methods as well.

An extension of psi research can also be made on the neurological side of the psychiatric field. There is, for instance, good reason for interest in the psi capacity of brain lesion cases and a still more imperative need to broaden ESP studies to include the use of neurological methods of obtaining responses in the tests. The earlier use of the galvanometric reflex by Dr. Brugmans has never been followed up, though his experiment indisputably justified repetition. Other measures, such as electroencephalography, that might serve to register the subtle psi reaction more sensitively than the usual verbal and manual methods, should be tried on an exploratory basis.

There is now a new stimulus to interest in the use of neurological techniques for measuring the subjects' responses in ESP tests. This is the finding to be reported by Dr. J. G. Pratt in an

early issue of the JOURNAL that the subject himself unconsciously registers success in the test. The sequence of responses shows that a hit affects the subject's choice of the next symbol. This means then that some subjects *do* make a reliable response to success. Perhaps this fact of success can be physiologically registered as well; if so, the whole subject of ESP will be removed from its state of uncertainty and dependence on statistical probability and will be put on a basis of reliable use.

* * * * *

Finally, what has the parapsychologist on his part to contribute to psychiatry? Because of the youth of the new field there is danger of our seeming to be presumptuous in claiming to be of any help to an older, better-established branch. The findings in parapsychology are all relatively new as yet and need more time for general acceptance. In addition to this, there is always a speculative element in stretching the implications of a new finding to its fullest apparent significance.

With these reservations, however, we can guardedly say a few things about the bearing of the psi researches on psychiatry. For one thing the methods of testing and appraisal developed in parapsychology could be useful in practical diagnosis. They have, in fact, already been given a small amount of application. Certainly the more we know of psi phenomena the easier should be the diagnosis of abnormal cases complicated by those types of experience (hallucinations, delusions, and such) commonly involved in spontaneous psi occurrences of a perfectly healthy order.

The question is frequently asked whether psi capacities themselves can be reliably used as an instrument of inquiry by the psychiatrist in diagnosis, or as a means of improving effectiveness in therapy. There are many psychiatrists who admit in confidence to taking seriously the possible operation of telepathy in effecting good rapport with the patient. There is still no clear certainty that telepathy does function reliably in that way, but if it does, as is entirely possible, there can be no doubt of the importance of such a finding. These problems are mainly research issues for the future.

In psychokinesis we have a psi-principle that seems pretty clearly applicable to psychiatry, to psychosomatic medicine in particular, though to what extent it applies remains to be discovered. The

psychogenetically induced stigmata—dermographia, blisters, and like effects—reported by medicine still stand unexplained. The mechanisms involved in other phenomena such as wart removal and the sudden whitening of hair from wholly psychical causes are yet to be found. It is possible PK will eventually be of some help in interpreting these organic occurrences which are at present inexplicable.

The major contribution of the psi studies to psychiatry, however, is one which, when it is accepted, may bring about basic changes in viewpoint and practice. It concerns the thought-brain relation. Most psychiatrists, like most psychologists, have settled down to the working assumption of a psychosomatic unity of some unknown sort. And since he knows more about the physiological end of this unity-axis, the psychiatrist has gradually drifted around to a physicalistic approach in dealing with the mind-body system of human personality, even though he readily admits he has no notion whatever as to how far physics really goes in explaining his patient's mental life. Since it does go a long way toward explaining neural action, and since the psychiatrist has modern science's horror of the confusion of dualism, he is driven toward a narrowly physicalistic explanation of the patient and his problem. Worst of all, he is discouraged from looking for anything nonphysical, since it would only complicate his conception of the individual.

This need for unity is understandable. With our conception of causation, some kind of unifying explanation will have to be found to account for all the interacting elements in nature. The pressing question is one, rather, of *how much diversity we have within this unity.* The parapsychological findings indicate that we shall have to look still farther back than we have hitherto supposed for a common basis of psychophysical interaction. They have shown that psi, transcending space and time as it seems to do, does not belong in the field of physics. Hence the experiencing system of the individual which we call the mind is in some degree nonphysical. This finding puts psychotherapy (as it does all fields dealing with personality) on a foundation of its own. Psychiatry is not just applied physics, a branch of engineering, as it would otherwise have to be.

This will mean for psychiatry the turning of some of its researches more directly toward discovery of the basic properties of

conscious experience, and what thinking and emotion, normal or abnormal, really are first of all in their own right. Failing, as we have been, to understand and explain them in terms of concomitant nerve physics, should we not strike out in a new direction? If we try looking at the problems of mental life in terms of unique principles and properties (searching, of course, for all there may be of these) we may discover how the personality works and how it gets out of order. If parapsychology is right in its evidence the turning to this new point of view is not an optional one. If it has opened the way to so significant an eventual shift of viewpoint for psychiatry, parapsychology will have done something in repayment.

J. B. R.

FURTHER WORK OF DR. STUART ON INTEREST TEST RATINGS AND ESP

By Betty M. Humphrey

ABSTRACT: The late Dr. C. E. Stuart had devised an interest inventory consisting of 60 items which subjects in ESP experiments were asked to check on a five-point scale varying from "like it very much" to "dislike it very much." The purpose of the inventory was to provide data on the basis of which it might be possible to separate the high-scoring ESP subjects from the low-scoring ones. By an analysis of the subjects' answers, Stuart devised a method of dividing subjects in a given series into two groups: the mid-range group and the extreme group. Prior to his death he had applied this interest measure to eight ESP card-test series. The results of the analysis are reported here together with those of eight additional ESP series to which the inventory measure was applied by Dr. Humphrey. The mid-range group in the eight series which Stuart himself had analyzed gave an average ESP score above "chance," although it was not significant. The results of the extreme group, however, were significantly below the expected chance average. The difference between the scores of the two groups was significant with odds against chance of over 700 to 1. The eight additional series which Dr. Humphrey analyzed showed the same trend as did the Stuart Section, but the results were not significant. When both sections were pooled, the difference between the mean ESP scores of the mid-range and extreme groups was such as would be expected from chance only once in about 714 times. Further analyses of data are discussed.—Ed.

Introduction

In 1945 and 1946, just prior to his death, Dr. C. E. Stuart was exploring the possibility that high-scoring and low-scoring subjects in ESP tests could be differentiated on the basis of their responses to an interest inventory. He devised an interest test consisting of a list of sixty items including activities, events, and objects familiar to college students. Subjects in a number of ESP experiments were asked to fill out the inventory by indicating their reaction to each item on a five-point scale varying from "like it very much" to "dislike it very much." By various analyses of the patterns of "likes" and "dislikes" of the subjects, Dr. Stuart tried to find some measure or measures that would separate the high-scoring ESP subjects from the low-scoring ones. At the time of his death in 1947 it appeared that he had found several such measures.

One of these measures was described in a report which Dr. Stuart himself published in 1946 (2). From a study of the sub-

jects' answers to the sixty items, he had devised a scoring method which separated the subjects into two groups: the "mid-range" group and the "extreme" group. The mid-range group consisted of those subjects whose inventory scores clustered about the median of all scores in a given series, while the extreme group was made up of subjects whose scores were at either the upper or lower end of the inventory score distribution. In the article referred to, Dr. Stuart reported that the subjects in the mid-range group obtained ESP scores above "chance" in a clairvoyance drawings experiment, while the extreme group subjects scored below "chance." The most striking result of the analysis, however, was that the subjects in the extreme group showed striking backward displacement; that is, they tended to hit the just-previous target picture rather than the intended one.

It seemed best, in undertaking to survey Dr. Stuart's unpublished research involving his interest test, to begin with a report of those series to which this same inventory measure had been applied. The available experiments consist of sixteen clairvoyance card-test series. To the first eight of these, Dr. Stuart himself had applied the interest inventory measure described above. In the remainder of this paper, these eight series will be referred to as the "Stuart Section," indicating that Dr. Stuart himself carried out the analysis of these data.

Late in 1946 and in 1947, eight additional series in which the interest inventory was used along with clairvoyance card tests, were carried out by experimenters at the Parapsychology Laboratory. To these series I applied the same interest inventory measure that Dr. Stuart had used in the first eight series. These later series will be known as the "Second Section" in subsequent discussion.

This report, then, will deal with the results of the Stuart interest measure as applied to sixteen experiments, all of which involve clairvoyance card tests.

Description of the Series Analyzed

Seven of the series in the Stuart Section were group ESP tests carried out with psychology classes by Dr. Burke M. Smith. (These series are referred to hereinafter with the letter "S" preceding the series number.) In one series (M1) Miss Elizabeth McMahan tested the subjects individually. In all series the procedure used

was "BT" in which the subjects recorded their respective calls for each card as it was held apart from the deck by the experimenter. The cards were of course screened from the subjects' view in all tests. Since the details of each experiment are not important in the present connection, no further discussion of procedure will be given. The main facts concerning the series are summarized in Table 1. Six of the eight series[1] have already been reported in detail elsewhere; the two previously unreported series (S5 and S6) were similar in conditions and procedure to the earlier series.

Table 1

GENERAL SUMMARY OF INFORMATION CONCERNING SERIES IN STUART SECTION

Series	Date	Experimenter	Number Subjects	Number Runs per Subject	Total Runs	Dev.	Reported (*Journal of Parapsychology*)
S1.....	5/46	B. M. Smith	34	4	136	− 7	1946, 10, 269-89
S2.....	5/46	B. M. Smith	32	4	126*	−20	1946, 10, 269-89
S3.....	5/46	B. M. Smith	25	4	98*	+13	1946, 10, 269-89
S4 Sm..	5/46	B. M. Smith	28	2	56	0	1947, 11, 118-46
S4 Ma..	5/46	B. M. Smith	34	2	68	+12	1947, 11, 118-46
S5.....	7/46	B. M. Smith	36	2	72	+ 6	not reported
S6.....	11/46	B. M. Smith	30	4	120	−58	not reported
M1....	11-12/45	E. McMahan	20	4	80	−10	1947, 11, 118-46
Total			239		756	−64	
					CR = 1.16		

*In Series S2 and S3 one subject completed only two runs instead of the usual four.

Altogether in the Stuart Section there were 239 subjects who gave a total of 756 standard runs of 25 trials each. There were 64 hits fewer than the total number expected from chance. The critical ratio is only 1.16, which, of course, does not approach significance.

In the Second Section, all eight series consisted of group clairvoyance tests. None of the series have been previously reported. Table 2 summarizes the main facts for each experiment. Two series were carried out by Dr. Smith, four by Dr. J. B. Rhine, and two by

[1] In the series known here as "S4," Smith tested a group made up of his own students and of the students of Dr. T. P. Malone. Although the entire group took the tests together, the results were analyzed separately for Smith students ("S4 Sm") and for Malone students ("S4 Ma") because of the possibility that a different degree of acquaintance with the experimenter might influence the results of the two groups. These series, along with M1, were reported together in 1947. In that report (as listed in Table 1) Series S4 was known as the "Group Series," while Series M1 was called the "Individual Series."

myself. (In Table 2 "S" preceding the series number indicates that Smith was the experimenter, "R" refers to Rhine, and "H" to Humphrey.) In three series the BT procedure was used as in the series of the Stuart Section. In four other series the procedure followed was "DT" in which a deck of cards, previously shuffled and enclosed in its box by an assistant, was placed before the group, and the subjects were asked to call "down through" the pack.

Table 2

GENERAL SUMMARY OF INFORMATION CONCERNING SERIES IN SECOND SECTION

Series	Date	Experimenter	Number Subjects	Number Runs per Subject and Procedure	Total Runs	Dev.
S7	12/46	Smith	40	2 Runs DT	80	+32
S8	12/46	Smith	29	2 Runs BT	58	− 6
R6	11/46	Rhine	37	2 Runs BT	74	+ 4
R7	12/46	Rhine	35	1 Run 5 x 5	35	− 9
R8-A	5/47	Rhine	36	4 Runs DT	142*	+ 1
R8-B	5/47	Rhine	40	4 Runs DT	160	+11
H2	12/46	Humphrey	28	2 Runs BT	56	−10
H4	12/46	Humphrey	33	2 Runs DT	66	+16
Total			278		671	+39
					CR = .75	

*One subject made two runs instead of the usual four runs.

For Series R7 a deck of cards was shuffled by an assistant (not present at the experiment), and each card was enclosed in a small opaque black envelope. These envelopes were then stapled onto a sheet of cardboard in a five-by-five arrangement—five rows and five columns. At the time of the experiment this board was placed in front of the subjects, who were asked to record their calls on a record sheet containing five rows and five columns also.

Since these series in the Second Section have not been reported before, it might be well to mention that in all precautionary details they did not differ from the reported series of the Stuart Section. For example, in all series the subjects made their calls in duplicate, the original copies in every case being collected and placed in the experimenter's possession before the true card order was revealed. And, as in all serious experimentation, the cards were always screened from the subjects' view.

In the eight series of the Second Section 278 subjects participated. The total of 671 runs gave a deviation of +39, which is not significantly different from "chance."

The sixteen experiments were not carried out with the sole objective of obtaining data for the interest inventory research; actually, the inventory was most often given incidentally in these series which were designed primarily for other purposes (such as to obtain additional information on other personality measures). Consequently, the psychological conditions of the series varied widely. Especially is this the case with the series in the Second Section, which are particularly heterogeneous in this regard. For example, in R8-A and R8-B, half the runs were made for cards whose location (in another building) was not known to the subjects. Nevertheless, in the present report, all clairvoyance card runs are considered together regardless of differences in conditions. This practice of course puts the greatest amount of strain on the discriminating power of the interest inventory measure.

In one series of the Stuart Section and in six series of the Second Section other types of ESP tests, such as GESP and precognition, were also given. The present report, however, deals solely with the clairvoyance card data; later reports will be devoted to the other types of ESP test.

The data from both sections have been completely rechecked both for direct hits and for backward displacement hits. Similarly, all computations have been rechecked.

The Interest Inventory Analysis

The Stuart Interest Inventory, which was published in the previous article on this test (**2**, p. 157), instructs the subjects to indicate after each of the 60 items one of the following answers:

LL if you *like it very much*
L if you *like it*
I if you are *indifferent* in interest
D if you *dislike it*
DD if you *dislike it very much*

In order to score the inventory, the subjects' responses are converted into numbers by giving +2 for every LL answered, +1 for every L, 0 for each I, −1 for each D, and −2 for each DD. The

algebraic sum of the numbers given to the 60 responses is the subject's interest inventory score.

For each experimental series considered alone, the interest scores are arranged in order of decreasing magnitude and the quartiles are found. The scores are then divided as closely as possible into two groups: (1) those in the second and third quarters, which constitute the "mid-range" group; and (2) those in the first and fourth quarters, which make up the "extreme group."

It is difficult with some series to secure an equal number of cases in each quarter. The practice followed here was to try to put approximately one-fourth of the total number of cases in the first and fourth quarters without separating the cases occurring at any given score level; for example, if four subjects obtained the same inventory score, all four cases were assigned to *either* the mid-range *or* the extreme group depending on which category had the fewer number of cases. They were never divided between the two groups. For this reason and also for the reason that some series had an odd number of cases, it was not always possible to have an equal number of subjects in the mid-range and extreme groups. When there was any doubt about a particular case, the "safest" disposition possible was made; that is, if the subject in question had a negative deviation, he was assigned to the mid-range group, whereas a case with a positive deviation was put in the extreme group since, according to the Stuart hypothesis, a positive deviation is expected for the mid-range group and a negative one for the extreme group.

The mid-range and extreme groups were computed separately for each individual series for the two sections reported here. Stuart had followed the practice of treating each series as a unit for this analysis because it seemed that the general conditions of each experiment and even the experimenter himself might exert some influence on the distribution of "likes" and "dislikes" registered by the subjects. An inspection of the median scores of the various series does reveal striking group differences. Whether these differences are due to the composition of the group, to the experimenter, or to the experimental conditions cannot be determined on the basis of present data.

As for other data reported here, all interest inventory scores and computations based thereon have been completely rechecked.

For each of the sixteen series here reported, the ESP scores for the subjects in the mid-range group were added together and were compared with the total ESP score for those in the extreme group. All evaluations of the data were based on the binomial formula.[2]

Results

Direct Hits

The subjects in the mid-range group for the eight series of the Stuart Section gave a total of 378 ESP runs with a positive deviation of 56, which is not significant. The 378 runs of the extreme group, however, gave 120 fewer hits than expected from chance. The critical ratio for this group alone is 3.09, which has a probability of .002. The difference between the mean ESP scores of the two groups is statistically significant with a CR of the difference of 3.18. As may be seen in Table 3, seven of the eight series show the trend given in the total, while in the eighth series the extreme group obtained two more hits than did the mid-range group. The total results are in the same direction as those reported by Stuart in his clairvoyance drawings experiment, although in the latter case the difference was not significant.

[2] Theoretically, when more than one subject makes calls for a single deck of cards, spurious results might be obtained because of certain "pattern preferences" on the part of the subjects. If, for example, there was a general tendency for subjects to prefer certain symbols or sequences of symbols, the variance would be affected in an unknown direction and the binomial formula would not be appropriate.

As has been pointed out before in connection with drawings experiments (**3**), when the significance of the data in question depends on a difference between the scores of two groups of subjects, it would be necessary to hypothesize that *different* pattern preferences were operating in the two groups of subjects to produce a spurious difference. Dr. T. N. E. Greville has devised a method of correctly evaluating such "multiple-calling" data which he published in the *Annals of Mathematical Statistics,* 1944, Volume 15, No. 4, pp. 432-34. This method has been modified to deal with the case of differences and has been applied in other connections to a number of the series involved in the present report. The method was applied to differences obtained between expansive and compressive subjects and to differences obtained when the subjects were separated into two groups by means of another interest inventory measure. In every case the CR of the difference obtained differed from that obtained by the binomial formula by only a few hundredths of a point. (For example, for Series S1, S2, and S3, CR's of the difference obtained by the binomial formula were 2.79, 2.81, 2.29. By Dr. Greville's method the comparable CR's of the difference were 2.76, 2.75, 2.26.)

Because of the almost prohibitive amount of work involved in carrying out the extensive analyses according to the Greville method, and especially because of the fact that no appreciable difference has ever been found in connection with any of the series of this report between the results obtained by the Greville method and the simple binomial formula, it was deemed unnecessary to apply the Greville method to the series of the present report where the CR's obtained were not of *borderline* significance.

Table 3

ESP Results for Mid-Range and Extreme Groups, Stuart Section

Series	Mid-Range			Extreme			CR_d
	Number Subjects	Runs	Dev.	Number Subjects	Runs	Dev.	
S1	17	68	+22	17	68	−29	2.19
S2	16*	62	− 1	16	64	−19	.81
S3	13*	50	+13	12	48	0	.64
S4 Ma	17	34	+ 9	17	34	+ 3	.36
S4 Sm	14	28	+14	14	28	−14	1.87
S5	18	36	+ 2	18	36	+ 4	(−) .12
S6	15	60	−15	15	60	−43	1.28
M1	10	40	+12	10	40	−22	1.90
Total	120	378	+56	119	378	−120	3.18
		CR=1.44			CR=3.09		

*In Series S2 and S3 one subject did two runs instead of the usual four.

In this section, then, the interest inventory analysis has revealed a significant effect in series whose total deviation was not different from "chance."

When the ESP results were compiled for the series in the Second Section, the same trend was found: the mid-range group gave a positive deviation while the extreme group gave a negative devia-

Table 4

ESP Results for Mid-Range and Extreme Groups, Second Section

Series	Mid-Range			Extreme		
	Number Subjects	Runs	Dev.	Number Subjects	Runs	Dev.
S7	20	40	+13	20	40	+19
S8	14	28	− 2	15	30	− 4
R6	19	38	+ 6	18	36	− 2
R7	17	17	− 5	18	18	− 4
R8-A	17	68	+ 1	19*	74	0
R8-B	20	80	+16	20	80	− 5
H2	14	28	+ 5	14	28	−15
H4	17	34	+18	16	32	− 2
Total	138	333	+52	140	338	−13
		CR=1.42			CR=.35	
			CR_d=1.26			

*One subject did two runs instead of the usual four runs.

tion. The total score of neither group is significant, however, nor is the difference significant. (See Table 4.) Only four of the eight series show the difference in the expected direction, while the other four series show practically no difference between mid-range and extreme groups.

When the results are pooled for the sixteen series, the following figures are found:

	Mid-Range			Extreme			CR_d
	Runs	Dev.	CR	Runs	Dev.	CR	
Stuart Section.....	378	+ 56	1.44	378	−120	3.09	3.18
Second Section.....	333	+ 52	1.42	338	− 13	.35	1.26
Total.....	711	+108	2.03	716	−133	2.49	3.19

The critical ratio for the mid-range group does not quite reach the .01 level of significance. The deviation of the extreme group is, however, significantly negative. The critical ratio of the difference between the mid-range and extreme groups is still highly significant (3.19).

Thus, when all series which were available at the time of preparing this report are pooled, we find that the measure derived from the interest inventory has, to a significant degree, separated high- and low-scoring subjects in clairvoyance card tests. From an inspection of the figures given in the preceding table, it can be seen that the measure was most successful in the case of the extreme group, the negative scoring subjects (CR = 2.49). Through an analysis based on the interest inventory, a significant difference has been revealed in series whose pooled total of 1,427 runs gave a negligible deviation of −25.

Backward Displacement

Since the previous Stuart report (2) on the clairvoyance drawings series showed a significant difference between the mid-range and extreme groups with regard to backward displacement, the calls of the subjects in the two sections here[3] were also checked for cor-

[3] Series R7 was not checked for backward displacement because the target cards were spread out in a five-by-five rectangular arrangement and the exact order of the subjects' calls was not known.

respondences with the preceding target cards, that is, for (−1) displacement.

Since the results for neither section are significant, the data for each series will not be given separately. The backward displacement figures for each section are:

	Mid-range		Extreme	
	Runs	Dev.	Runs	Dev.
Stuart Section	362.88	−47.40	362.88	+8.60
Second Section	303.36	−47.80	307.20	+1.00
Total	666.24	−95.20	670.08	+9.60

Although the mid-range group shows a tendency toward avoidance of backward displacement as did the mid-range group in the Stuart report, the deviation does not approach significance (CR = 1.84). Whereas in the Stuart clairvoyance drawings series the extreme group had given a significantly positive deviation, the extreme group here in the card data shows only a negligible positive deviation. Only one of the fifteen series shows any marked tendency in the direction indicated in the previous Stuart report: the extreme group in Series S2 gave a CR of 2.86 (from a positive deviation), and the CR of the difference between the two groups for this series was 2.59. Since this is only one series selected from many, we can only conclude that the inventory measure has failed to reveal significance with regard to backward displacement in these ESP card series.

Discussion

The Stuart interest inventory measure proved to be capable of separating the high- and low-scorers in sixteen clairvoyance card-test series. In spite of the heterogeneity of conditions involved in the sixteen series, a CR of the difference between mid-range and extreme groups has been obtained which is not easily ascribable to chance (CR_d = 3.19). Those subjects whose interest inventory scores did not differ markedly from the group median gave positive ESP results as a whole, while that group of subjects whose scores were at the extremes of the interest score frequency distribution (either high or low) gave significantly negative ESP scores.

The finding that the Stuart inventory measure has separated subjects into two groups whose mean ESP scores are significantly different is a valuable contribution in the search for more insight

into the nature of ESP. We gain increased control over the phenomenon by knowing, for example, that those subjects whose interest scores fall in the extremes of the score range for the group would on the whole be expected to give negative ESP scores. Although this finding is in itself an important one, it raises a further question: whether the positive score of the mid-range group and the negative score of the extreme group were due primarily to the large deviations of only a few subjects, or whether a majority of the subjects in each group contributed to the total score. In other words we want to find out about the consistency of trend within the mid-range and the extreme groups.

For the Stuart Section, we find that, in the mid-range group, 52% of the subjects obtained positive ESP scores, 8% gave a deviation of zero, while 40% showed negative deviations. In the extreme group the comparable figures are: positive, 35%; zero, 8%; negative, 57%. The relationship of the inventory measure and ESP scores may be further evaluated if the figures from which the above percentages were derived are placed in a 2×2 contingency table. In this analysis a count was made of the number of subjects in the mid-range and extreme groups who gave a positive deviation on the ESP tests (regardless of the magnitude of the deviation), and the number of subjects whose deviation was negative or zero. Following the practice of both Stuart and myself in other research, we arbitrarily assigned to the negative category those subjects who gave zero deviations. The following fourfold table resulted:

STUART SECTION

	+ESP	−ESP	Total
Mid-range Subjects	62	58	120
Extreme Subjects	41	78	119
Total	103	136	239

The occurrence of the larger number of cases in the upper left and lower right cells indicates a relationship between positive ESP scores and mid-range inventory scores on the one hand, and negative ESP scores and extreme inventory scores on the other. The

chi-square derived from this table is 7.22 (1 d.f.), which has a probability of less than .01. This evaluation shows that it is improbable that ESP scores and inventory scores are independent; in other words, positive ESP scores appear to be associated with mid-range inventory scores, and negative ESP scores, with extreme inventory scores. The figures in the fourfold table as well as the percentages given earlier, show that the findings cited in the Results Section are not due to the presence of a few high-scoring subjects in the mid-range group and to that of a few very low-scoring subjects in the extreme group; the majority of the subjects in each group followed the trend shown in the total results for each group. Especially was this the case in the extreme group.

When the same analysis was made for the Second Section, the following fourfold table was found:

SECOND SECTION

	+ESP	−ESP	Total
Mid-range Subjects	65	73	138
Extreme Subjects	59	81	140
Total	124	154	278

Although the extreme group follows the trend shown in the Stuart Section (that is, more cases with negative deviations), the mid-range group does not follow the expected trend. The chi-square (.69) is not significant.

When both sections were pooled, the following distribution was obtained:

BOTH SECTIONS

	+ESP	−ESP	Total
Mid-range Subjects	127	131	258
Extreme Subjects	100	159	259
Total	227	290	517

The chi-square for this table is 5.91 (1 d.f.), which has a probability of less than .02. For both sections pooled, the actual percentages of subjects in the mid-range group who gave positive, zero, and negative deviations are: 49%, 11%, and 40%, respectively. Comparable figures for the subjects in the extreme group are: positive, 39%; zero, 12%; negative, 49%.

In his report on the interest inventory, Stuart adopted the term "affectability" to refer to the measure used here. The subjects in the extreme group in the one series he reported were those who registered a relatively large number of likes (LL's and L's) or dislikes (D's and DD's). Therefore, he termed this group the "affectable" group, while those who were average or moderate in their interests were called "unaffectable." Although the sixteen series in the present report were analyzed according to the method described by Stuart in his article, further study of the inventory scores throws some doubt on the correctness of the use of the term "affectable" in connection with the extreme group. Because of the fact that only a small proportion of D's and DD's were given by the subjects in these series, the inventory scores were all positive, ranging from +3 to +88. When the majority of the scores are positive, it is possible for a score at the lower end of the score distribution to be the result of an unusually large number of "indifferents" (I's), for which no weight is given in scoring. An analysis of some of the sixteen card-test series did reveal a number of highly "indifferent" subjects in the extreme group. A person who is in general indifferent could not properly be characterized as "affectable." For this reason, the term "affectable" has for the present been dropped in reference to the extreme group. Until a more detailed analysis of all series is completed, it seems best to withhold interpretations of the measure.

The reason for the failure of the Second Section to give significant results cannot be determined from the data of these series. The psychological conditions of the Stuart Section and of the Second Section were undoubtedly different, but it cannot be said with any feeling of certainty wherein the essential differences lay. It was pointed out earlier that the conditions of the series in the Second Section were heterogeneous; different procedures were used, special instructions to the subjects were given in some series, and the series

were carried out for various purposes. On the other hand, the first seven series of the Stuart Section were as similar in purpose, instructions, and procedure as possible—even the experimenter was the same. The eighth series (M1), however, differed from the other series in many ways, and yet that series yielded the largest difference between mean ESP scores of the mid-range and extreme groups. The many uncontrolled variables in both sections of the data make it impossible to reach any conclusions regarding the failure of the Second Section.

The fact that Series M1 did yield a large difference raises the possibility that the inventory measure may be more effective in series in which the subjects are tested individually. Whether or not the effectiveness of the interest inventory measure differs in individual and group tests will have to be determined by further research. At present it seems more plausible to suppose that other factors in the situation, such as the subject-experimenter relationship, will prove to be more important in determining the success with which the measure separates high- and low-scoring subjects.

Whether or not the interest inventory measure will prove effective in separating high- and low-scoring subjects in other types of ESP tests remains to be seen. The unpublished series of general ESP and of precognition tests are not yet extensive enough to permit of any conclusions.

In view of the great variations found in the performance of most subjects in ESP tests, the question naturally arises as to the consistency of the subject's ratings on the interest inventory. It is conceivable that the subject's ratings might be affected by the circumstances under which the inventory is filled out, by the manner and instructions of the experimenter, or by other factors, both psychological and physical. Whether or not subjects do obtain different ratings on inventories taken under varying circumstances cannot be determined on the basis of present data. We shall have to wait for research designed specifically to get at this problem.

The effectiveness of a combination of the inventory measure and other personality measures in separating "good" ESP subjects from "poor" ones has already been studied to some extent. The expansion-compression measure based on the form quality of subjects' drawings has been combined with the inventory measure for the

data of the sixteen series in this paper, and a report on that research will be made in the near future.

One of the principal advantages of the interest inventory is the small amount of time involved in administering it. Subjects seldom require more than three minutes to record their responses to the sixty items. Thus it is easy to include the inventory in any experimental session, no matter what the major purpose of the research. Moreover, the scoring of the inventory is simple and quick and does not entail the great investment of time on the part of the experimenter that such personality tests as the Rorschach and Bernreuter do.

The full extent of Dr. Stuart's great contribution to our knowledge of the relationship of ESP and personality cannot possibly be assessed at the present time. The research reported here is only one facet of the work he was carrying out at the time of his death. The findings reported in this paper and those of his still unpublished work open up new areas which may take years to explore fully.

References

1. Stuart, C. E. An analysis to determine a test predictive of extra-chance scoring in card-calling tests. *J. Parapsychol.*, 1941, **5**, 99-137.
2. ———. An interest inventory relation to ESP scores. *J. Parapsychol.*, 1946, **10**, 154-61.
3. Stuart, C. E., Humphrey, B. M., Smith, B. M., and McMahan, E. Personality measurements and ESP tests with cards and drawings. *J. Parapsychol.*, 1947, **11**, 118-46.

Parapsychology Laboratory
Duke University
Durham, North Carolina

"DIRECT INFLUENCE" OF A PERSON UPON THE BEHAVIOR OF ANIMALS[1]

By W. Bechterev

The question of a direct mental influence upon animals merits special consideration, and for a long time I watched for an opportunity to deal with this question experimentally. A few years ago, shortly before the beginning of the Great War, I got such an opportunity on the following occasion:

Quite incidentally, after having been entirely indifferent to circus shows for several years, I visited the Cirque Moderne in Petersburg. On that evening, W. Durow showed his trained animals, among which was a St. Bernard dog who could count to nine. The dog, called Lord, seemed to be of a very quiet and reliable character.

The demonstrations were conducted in the following way: Durow asked someone in the audience to write down any numbers, the sum of which should not exceed nine, since, according to his explanation, Lord was not able to count accurately any higher. Two or three figures to be added up were written on a piece of paper or a slate by someone, and then shown to Durow, who had his back to the dog. Immediately after Durow's verbal command, the dog gave the sum of the numbers by barking. The experiments were repeated very often and were always entirely successful; the dog barked loudly and exactly the number of times indicated by the sum.

Durow chanced to notice my presence in the circus. Approaching me he said he was glad to meet me and asked for my assistance at a mutual elaboration of some experiments of his. We agreed to try some experiments in my apartment. On the appointed day

[1] The tests of telepathy in dogs carried out by the Russian neurophysiologist, Prof. W. Bechterev, have been mentioned in the literature of parapsychology, but a full account has not until recently been available in English. Utilizing a translation made by Dr. Gerda Walther, we present this abbreviated version of the original report which appeared in 1924 in the *Zeitschrift für Psychotherapie*. In addition to its historic interest the work of this distinguished scientist offers some valuable suggestions for a still largely unexplored area of problems. We have included the principal experimental conditions and results along with the author's interpretation and his suggestions for further research.

Prof. Bechterev, at the time the experiments were made, was President of the Psycho-neurological Academy and Director of the Institute for Investigation of the Brain in St. Petersburg.—Ed.

Durow brought two dogs to me, the above-mentioned Lord and a little fox terrier, Pikki. First I shall speak about the experiments with Lord.

The dog was placed on a couch where he sat quietly. Then Durow suggested that several figures be written on a piece of paper, the sum of which should not exceed nine. Subtractions could also be used, but again the difference was limited to nine. These numbers were shown to Durow, and with his back to the dog he immediately gave the signal, "Now, Lord, count." Lord barked as often as indicated by the sum or the difference of the two figures. These experiments were repeated many times, always with the same success.

I must point out that it means nothing to the dog if the numbers are shown to him, for he is not capable of counting and adding them up according to the written figures, as could easily be seen by experimentation. If Lord successfully gives the number indicated by the sum or difference of two figures, this only occurs because the dog gets an appropriate mental suggestion after the words: "Now, Lord, count." If such a suggestion is omitted the experiments are not successful. Three experiments which were done to throw light on this aspect of the problem had the following results: Although the single figures with a sign of addition between them produced fits of barking in the dog, the number of barks in no case was the same as the sum. When I tried to accompany the written figures with appropriate motions, the correct result was not obtained either, while the experiments with the so-called mental suggestion directed towards making the dog give the correct number of barks were always conducted successfully. Accordingly it is clear that only the concentration of the experimenter upon a consecutive series of barks from one to nine led to correct performance.

A series of similar experiments were conducted in my dwelling with the little fox terrier, Pikki, a very lively and nimble dog. Durow himself describes[2] his method in these and similar experiments in the following manner:

> Suppose we have the following task: to suggest that the dog go to a table and fetch a book lying upon it. I call him and he comes. I take his head between my hands as if I am symbolically inculcating in him the thought that he is entirely in my power. . . . I fix my

[2] In "My Quadruped and Winged Friends."

> eyes upon his. . . . I pull together all my nerve power and concentrate so that I entirely forget the outer world, impressing upon myself mentally the outlines of the object in which I am interested (in this case, the table and book) to such an extent that even when I look away it stands before me as though real. In the course of about half a minute I literally devour the object with my eyes, think of its minutest details. . . . Enough! I have memorized them.
>
> I turn the dog toward myself with an imperious gesture and look into his eyes, somewhere into his interior. I fix into his brain what I just before fixed in my own. I mentally put before him the part of the floor leading to the table, then the legs of the table, then the tablecloth, and finally the book. The dog already begins to get nervous, to become restless, tries to get loose. Then I mentally give him the command, or rather the mental push: "Go!" He tears himself away like an automaton, approaches the table, seizes the book with his teeth. The task is done.

The experiments were carried through in the afternoon in the presence of my family and some acquaintances. Altogether, six experiments were made, the first four by Durow and the remaining two by myself. One of these was as follows:

After the procedure of "suggestion," the dog was to jump upon one of the chairs at the wall behind Durow, then to climb on the little round table beside it, and, stretching himself, scratch the big portrait hanging on the wall above the table. This seemed a complicated act and not easy to carry out by a dog. But all the same after the usual procedure of concentration and looking into the eyes, Pikki, after a few seconds, jumped from his chair, quickly ran to the one at the wall, then equally quickly jumped upon the little round table. Rising on his hind legs, he reached the portrait with his right paw, scratching it a little with his claws. This task was known only to Durow and myself. I stood beside Durow, continually watching him as well as the dog without being able to notice anything which could explain how the task was fulfilled. In order entirely to convince myself I decided to make similar experiments myself in which I did not tell *anybody* of what I was thinking.

My task was that the dog was to jump upon a round chair about twelve feet away from me near the grand piano and sit down and remain there. As in the former experiments, the dog was asked to jump on the chair. Concentrating upon the round chair for some time, I gazed into the eyes of the dog, after which he ran away head over heels and raced around the dining table several times. I thought

it was a failure, but then remembered that I had exclusively concentrated my mind upon the form of the round chair, forgetting to begin with concentrating upon the movement of the dog toward the round table and then on his jumping upon the chair. Consequently I decided to repeat the same experiment without telling anybody what the task was or mentioning my mistake, but only correcting myself in the above direction. Again I had the dog sit down on the chair. I clasped both my hands around his muzzle and then began to think that he was to run to the round chair about seven and one-half feet behind me, jump upon it, sit down, and remain there. After a concentration of about one-half to three-quarters of a minute I let him go and I hardly had time to look around when he was already sitting on the round chair. The task performed in this case by Pikki as already mentioned was not known to anybody beside myself as I had not taken the advice of anybody in this direction. All the same Pikki had found out my secret without the least difficulty.

Durow left Petersburg the next day and the continuation of the experiments which we had planned after his return could not be carried out because of the commencement of the Great War. After the war I visited Durow in Moscow where I again made some experiments with Pikki. I did not see Lord again. He had died from a disease from which he had been suffering for a long time. Pikki was as well as ever, the same lively, brisk dog as before. I was of course interested in conducting some experiments of thought suggestion with him myself. This I succeeded in doing on two different journeys to Moscow.

During the first I personally was able to make five experiments according to the same method and with tasks similar to those of the first experiments. The task, however, was changed each time and was known only to myself. All of the five experiments, two of which were carried out in Durow's presence and three in his absence, must be looked upon as successful.

After having again convinced myself of the success of such experiments I decided to try some more on my next visit to Moscow, but this time introducing various test conditions which might lead to an explanation of the mechanism through which success is obtained in such experiments.

This time I did the experiments with Durow in the presence of Dr. S., one of my collaborators. One of these experiments, carried out by Durow, consisted of a task selected by me. The dog was to rush at a little stuffed wolf in the right corner of the room. The experiment was made in the usual manner. The dog was put upon the chair, his muzzle was taken with both hands, and he was looked at sharply for about half a minute. When let go he rushed at the wolf with such a fury, as if he wanted to tear it to pieces, that it had to be taken away from him.

Another of this group of experiments was conducted by me. The task, which I had not communicated to anybody, was that the dog was to jump on a chair and seize the towel lying on its arm. The usual method of suggestion by concentration was followed. After I had let loose his head, the dog immediately rushed to the intended chair; then, however, he turned back to the wolf and began to bark at it and rush at it with his former fury.

It was evident that the earlier suggestion, being of an emotional character, had left such a deep impression that it had not abated sufficiently before the later suggestion. And indeed it turned out that the next two experiments with thought suggestion also had the same result; that is, instead of fulfilling the suggestion, the dog turned to the wolf and began to bark at it furiously.

The following experiment also was done by me. The task was that the dog should pull a pocket handkerchief out of the right hand of Dr. S., who was standing a little distance back. The task was known only to me. The suggestion as usual did not take more than half a minute. After that the dog immediately rushed to the right hand of Dr. S. and tore the handkerchief from it.

We thought that in the earlier experiment involving the suggestion to rush at the wolf, the dog had reacted to the expression of Durow's face during the suggestion. Therefore it was agreed that this experiment should be repeated in such a way that Durow should either laugh or at least pretend to laugh during the suggestion. This Durow did.

Constantly under our close observation during the suggestion, Durow put on a laughing expression and, as before, never moved his lips. Nevertheless, after the suggestion the dog rushed at the wolf with the same fury as before, barking loudly. This time we

thought the dog might be directed by the movements of the experimenter's eyes during the suggestion. Accordingly Durow was asked to repeat the experiment in which the pocket handkerchief was torn out of Dr. S.'s right hand. However, this time when he made his suggestion his eyes were tightly blindfolded with a towel. At the same time we went into the adjoining room. With the usual procedure of holding the dog's head, Durow concentrated upon having the dog run to Dr. S. and take the handkerchief from his hand. This experiment was a complete success.

I have no special explanations to make about the above experiments. They are so striking in themselves that they merit attention apart from any explanations. But I must say that I am no longer astonished after the above experiments at stories of wild animals being banished by a look. The question is: what should be said about these experiments and how are they to be understood? First of all I must take it for granted that the dog was not influenced by the conversation about the tasks, since all experiments which I conducted myself were made without any conversation about any task and without informing Durow or anyone else what was to be done. In addition, all conversations were carried out with special care. Everything that could possibly have allowed the dog to be led by any verbal hints was omitted by us.

After the first series of experiments I thought there was only one possible explanation, which was that Pikki distinguished himself by an excellent memory. Accordingly one could surmise that the experimenter, while looking into the eyes of the dog, involuntarily moved the axes of his eyes (toward the object being concentrated on) and the dog observed it. The dog, already being accustomed to obey, immediately gave some signs of restlessness under the experimenter's gaze and, after being released, at once carried out the task. In this connection it is interesting to observe how Pikki at the conclusion of the suggestion jumps from the chair, head over heels, with all signs of excitement in his behavior until the task is done, after which he immediately calms down. My experiments with Lord could be explained in the same way. It can be supposed that the mental concentration upon a certain task is connected with eye or head movements of which the experimenter is not aware and which the persons about him do not notice without specially watch-

ing for them, while the trained dog, eager to perform, could easily notice them.

This explanation, however, does not tell why Pikki was entirely successful in carrying out his task when the experimenter's eyes were blindfolded. In this case he had no way of making use of such signs in the form of eye movements. The assumption must also be dropped that in the experiment involving the suggestion to rush at the stuffed wolf, the dog was directed by the expression of the experimenter's face, because the same experiment was later repeated successfully with Durow purposely distorting his face by an artificial smile. That all my experiments involved the carrying out of tasks known only to myself, some in the absence of Durow and other onlookers, must also be taken into consideration by the skeptics.

* * * * *

I shall now proceed to report further investigations carried out[3] with the trained dog, Pikki, by the medical doctors K. and P., who were sent to Moscow for that purpose. I am able to give a description of the doctors' experimental procedures, which were according to my proposal and plan. The conditions of the experiments were the following: the experimenter, Durow, had no contact with the dog before any of the experiments; he did not even see him. Durow entered the room only after the conditions of the experiment had been decided.

Experiment 1. This was a simple experiment in which the dog was to jump upon a chair, then upon a table and take the paper lying there. The experimenter had not touched the object in advance. Immediately after the "fixation" (that is, the mental suggestion) and the corresponding holding of the dog's head, the experimenter stepped to the side thereby giving a free path to the dog. During the experiment he had his back turned to the table. The dog, after the suggestion, jumped on the chair, then on the table, and took the designated paper, although a brush and a towel were lying beside it.

Experiment 2. The experimenter was supposed only to "fix" his eyes upon the dog while the assistant held the animal's head motionless. After a few minutes, the dog jumped upon a chair,

[3] On January 19, 1920.

then the sideboard, and took from it a certain spoon which he carried to the experimenter, all as planned. The experimenter followed the dog with his eyes all the time until the task was done.

Experiment 3. The experimenter, with blindfolded eyes, held the head of the dog. The task was for the dog to jump upon the knees of the doctor sitting near by and take the glove which was placed on them. After two minutes, the dog did so.

Experiment 5. The face of the experimenter, who held the dog's head, was hidden from the animal by a wooden screen. The dog was to go into another room, jump on the knees of the doctor there, and kiss him. The first attempt ended with a failure, probably because the chair upon which the dog was sitting was upset. The second attempt was also a failure; the dog carried out, not the task being thought of, but a task that had been suggested prior to the experiment in question but which was dropped by Durow's advice. It was to take a piece of bread out of a tube lying on the table. At the third attempt the dog quickly did what was desired of him.

Experiment 6. The face of the experimenter was hidden from the dog behind a metal screen; the head of the dog was "fixed" by the experimenter's hands. The dog was to jump upon a table, then a chair, and take a pear hanging above the table. After several unsuccessful attempts, including one attempt to seize the pear, the screen was taken away; but even so it was only after the assistant had pointed at the pear that the dog finally took it.

Experiment 7. This experiment was similar to the previous one except that the screen was of paraffin. The task consisted in having the dog jump upon the chair, then the table, and take a piece of paper. After three attempts, which were accompanied by passing the hands before the dog's face, the task was successfully completed.

Experiment 8. This trial brought positive results, but without any passes being made or the dog being touched. The task was for the dog to take an object from a chair. When a paraffin screen was used, there was no success. When it was replaced by a metal one, the experiment was successful. The experimenter did not touch the dog, but aroused his attention by calling his name while "fixing" his eyes through the screen.

* * * * *

I now come to a discussion of the experiments of Dr. F. who

was later sent to Moscow at my proposal. A total of nine experiments were made. Durow was the experimenter in the first of them, Dr. F. assisting; in the second, Dr. F. was experimenter, Durow assisting. The other seven experiments were carried out independently by Dr. F. in the absence of Durow who did not even know that work with the dog was planned. All these experiments can be called successful with the exception of the fourth one in which the dog was to bark at a rabbit in a cage. In this case, after a suggestion lasting seven minutes, the dog became excited and ran through the room past the cage, but did not carry out the task. It transpired that he had been strictly forbidden to bark at the animals in the cages. The experimenter asked Durow to repeat this experiment in his stead. Durow continued the suggestion, and after two minutes the dog fulfilled the thought, barking at the rabbit from a distance.

Further experiments in this group proved that neither the wooden screen covering the whole head of the experimenter from all sides, nor the paraffin screen consisting of "spectacles" thickly filled up with paraffin and worn over the eyes, prevented the execution of the thought suggestion.

To exclude the assumption that in carrying out a certain task the dog was guided by some involuntary sign given by the experimenter, I proposed an additional series of experiments in which the owner of the dog should not be present and the dog (Pikki) should not be led into the room until immediately before the experiment so that the experimenter might conceal himself either behind the screen or in some other manner. Only the assistant, who did not know the experimenter's task, should have contact with the dog during the experiment. Unfortunately the dog was ill at the time that these trials were possible. Nevertheless the following experiments, described by the assistant, are worth mentioning:

> Durow was absent. In the room of W. I wrote down the task: the dog was to seize with his teeth a paper ball lying under the table in A's room. Dr. F., who did not know the task, called the dog into A's room and put him on the chair, fixing his head. I was standing on the threshold, the dog about 28.5 inches away from me. At the first attempt of the animal to free himself I withdrew into W's room quickly, banging the door. According to Dr. F.'s account, Pikki, after jumping down from the chair, ran to the paper ball lying

under the table designated by me, touched it with his muzzle, then ran to and from two other balls doing the same. The other experiment was as follows: I concentrated on the dog's running from A's room to W's room and jumping on the easy chair there. The dog rushed into the other room past Dr. F., who did not know the task and who closed the door behind him. According to his report, Pikki executed the task essentially as directed.

A consideration of all the experiments carried out, those done according to my directions as well as my own tests with the two trained dogs, enables us to draw the following conclusions:

1. The behavior of animals, especially that of dogs trained to obey, may be directly influenced by thought suggestion.

2. This influence may be effective without any direct contact between the sender and the dog who is receiving, as when they are separated by a wooden or metal screen, blindfolds, paraffin-filled "spectacles," etc., and even when the experimenter does not himself know the thought suggested to the dog.

3. From this it follows that the dog may be directly influenced without the presence of any signs by which he could be guided.

It is hardly necessary to say how desirable further experiments in different directions with dogs would be. In these further experiments it would be important to study, not only the conditions governing the transfer of the mental influence from the agent to the percipient, but also the circumstances involved in both the inhibition and the execution of such suggestions. This would necessarily be of theoretical as well as practical interest.

I have started such experiments myself with my young dog, Gabish, which was entirely untrained in this direction. Regarding these experiments with Gabish I will mention that after some practice he began to carry out the mental command; for example, he jumped upon and sat down on a chair of which I had thought. Unfortunately these experiments had to be interrupted because the dog, after missing a leap, sprained his shoulder and began to limp badly.

This attempt has shown that the above experiments can be done not only with trained but sometimes also with untrained animals after some practice, of course, to accustom them to the experiments.

In this connection I suggest that during the period of accustoming the dogs to such experiments one should not at the very start impose rigid conditions excluding any possibility that the dog may

make use of some slight sign given unconsciously by the experimenter (as for example, movements of the eyes). It is desirable to avoid putting the object to be sought in an unusual place, because of the danger of discouraging the animal from the very beginning. Only after the dog has succeeded to some degree in going to the intended objects and has been encouraged with food or caresses, should one proceed gradually to experiments in which every possible opportunity to make use of outer signs can be entirely excluded.

A MECHANICAL SCORING TECHNIQUE FOR TESTING GESP

By S. David Kahn and Ulric Neisser

ABSTRACT: Two series of ESP tests were carried out by means of a procedure that permitted checking of the results on an automatic test-scoring machine. In the first series, 21 subjects at Harvard University were provided with IBM answer sheets containing space for 300 trials. On each trial the subjects were to mark one of five possible choices in the attempt to duplicate a marked target sheet which had been locked in a file immediately after it was prepared. The second series, in which 63 subjects took part, differed from the first chiefly in two respects: the target series was only half as long; and it was a distance experiment, with the target sheet being located at Harvard, while the subjects were 500 miles away. Positive scoring was obtained in both series, and the combined results were statistically significant.

This experiment was performed by the Harvard Society for Parapsychology, of which Mr. Kahn is President and Mr. Neisser is Director of Research.—Ed.

Introduction

In recent years some critics have questioned the validity of the evidence for extrasensory perception on the ground of the possibility of personal errors in scoring. They point out that the evidence consists principally of small deviations from an expected chance mean and that even slight scoring errors might go far toward accounting for these deviations. Some have even maintained that, because of the revolutionary concepts involved in the notion of ESP, accidental or even unconsciously motivated errors are a more plausible explanation of the results obtained, and that until the bare theoretical possibility of such errors has been eliminated, the evidence for ESP cannot be considered conclusive.

To a considerable extent these criticisms have already been met in work done to date. First and foremost, the method of independent checking of all results by at least two experimenters has been universally adopted. It is highly improbable that under such circumstances any significant amount of error could remain undetected. Second, salience effects have been discovered, and hits are generally found to be distributed in a definite nonrandom fashion down or across the page. Analyses of distribution were made of

data collected before the discovery of these effects, and significant results were obtained. It seems highly improbable that such a consistent distribution could occur through error, either accidental or unconsciously motivated.

Nevertheless, for theoretical reasons, it was felt that it would be desirable to find a technique of scoring in which the possibility of human error would be more completely eliminated. A mechanical scoring device is one obvious solution to the problem. The International Test Scoring Machine of the International Business Machines (IBM) Corporation was therefore selected, and a convenient, practical technique of ESP testing and scoring was developed with the use of this machine.

Two experimental series have been completed. The first, designed simply to test the practicability of the use of the IBM machine technique, was run at Harvard in May, 1949. The second series, in June, 1949, was a long-distance GESP test in which the subjects were at Virginia Beach, Virginia, and the target at Harvard, a distance of about 500 air miles. In this series a preliminary investigation of the effect of task set, or attitude toward the test, was also attempted.

Methods

In the first series, the standard "answer sheet" IBM form I.T.S. 1100 B 107 was used both as target and record sheet. On each side of this answer sheet are 150 calls or sets of five alternative choices. The sets are numbered 1-300 and are arranged in columns of 30. Each set consists of five small vertical "rectangles" arranged in a horizontal row, as illustrated.

One of these sheets was prepared as the target. One rectangle on each of the 300 sets was blacked in, using Tippett's tables of random numbers. Thus the target sheet consisted of 300 individual targets, randomly arranged. No effort was made to insure that an equal number of targets would appear in each position (rectangles 1 through 5). The experimenter who prepared the target sheet had no contact with any of the subjects. The target sheet was placed

in a locked filing cabinet immediately after preparation, and no other person had access to it until the conclusion of the series.

In the first series, 21 subjects, mostly students, were used. These were selected from a group of about 45, all of whom had been previously tested with ESP cards. The 21 chosen were those who had averaged 5.3 or better with the cards on 10 or more runs. It was hoped that among such a selected group there would be more positively scoring subjects than would be found in a random sample.

Each subject was given an answer sheet and was informed of the nature of the target. He was told its general location; that it was in a filing cabinet in the Harvard Psychological Laboratories, Memorial Hall, Cambridge, Mass. He was then asked to duplicate this target as nearly as possible; that is, to make 300 successive guesses, each among five alternatives. He was to do this at his leisure, at any time within a period of several days.

In the second series the "answer sheet" IBM Form I.T.S. 1000 B 106 was used. This sheet differs from the one used in the first series in that it has space for only 150 calls, the back of the sheet being blank.

A target was again prepared using Tippett's tables. This time, however, an equal number of rectangles (30) were filled in for each of the five possible target positions. Within each half of the target series (that is, from targets 1 to 75, and 76 to 150), the five target positions were equalized in the following manner: numbers 1 and 2 in Tippett's tables were designated to represent the first target position; 3 and 4, the second; and so on for all five positions. Following an arbitrary method of taking the figures from the tables, the experimenter translated Tippett's numbers into the five target positions until a given position had been represented 15 times (1/5 of the 75 trials in each half of the target page). The Tippett numbers pertaining to this position were then disregarded. When another target position was represented 15 times, all Tippett numbers referring to it were then ignored, and so on, until each target appeared 15 times in a given half of the target record.[1]

[1] This method of preparing a closed target sequence did not produce a random order of events because of the restrictions imposed toward the end of each half of the target sheet. As the analyses usually applied in such situations are based upon the assumption of "randomness" throughout the target series, the method raised questions that could be answered only by further study of the data. If subjects showed preferences for certain response positions, these might operate to

The target sheet was prepared in Cambridge, where it remained until the conclusion of the tests. The experimenter, S.D.K., who administered the test in Virginia, left Cambridge before the preparation of the target. There was therefore no possibility of any sort of sensory cues.

Sixty-three persons, largely adults, all of whom believed in the theoretical possibility of scoring more than the chance expectancy of hits by the use of extrasensory perception (and who were thus all "sheep" according to Schmeidler's terminology) were subjects in the second series. The target was hung in S.D.K.'s locked room at Harvard in a sealed opaque envelope. The subjects were shown a photograph of the building and of the room with the envelope in position. (The target was not in the envelope at the time the picture was taken.) The subjects filled in their sheets at leisure.

An attempt was made to establish a task set in the subjects by a simple verbal method. Before filling out his sheet, each subject was asked to answer yes or no to the following question: "Do you *personally expect* to get more correct hits than you would expect to get by chance alone?" The experimenter insisted that the subjects commit themselves, even those who claimed that they had no expectations one way or the other.

At the conclusion of each series the results were scored on the IBM machine. In the first series this was done by the Harvard Office of Tests; in the second, by the Educational Records Bureau in New York. Each subject's sheet was scored for total hits, and the possible occurrence of a decline effect was studied by comparing the totals of the first half of each record sheet with those of the second half.

---

give spuriously high or low scores in the nonrandom part of the target sheet.

In a recheck at Duke University it was found that the second, third, and fourth positions were called by the subjects much more often than the first and fifth. In view of the unusual way in which the target order was made up, it seemed safer to evaluate the truly random sections of the target sheet separately. The targets in the first half of the record were random from the first through the fifty-third, and in the second half they were random from the seventy-sixth through the one hundred thirty-fifth. The random sections of the data (trials 1 to 53, and 76 to 135) were re-evaluated by the Greville method and a CR of 2.24 was obtained. Thus it appears evident that in this particular series the nonrandom character of part of the target sequence did not produce spuriously high results. The CR of 1.74 in Table 3 may therefore be considered to be on the "safe" side in evaluating this second series.—Ed.

Results

The machine-scoring method used is based on the assumption that each subject marked out one, and only one, choice in each set on the answer sheet. In the recheck in the Parapsychology Laboratory at Duke University it was found, however, that subjects tended to leave some sets completely unmarked, and occasionally they marked two choices in one set by mistake. In Series I there were 11 sets left unmarked and 11 in each of which two marks were made. These sets were omitted and the results obtained by machine scoring were corrected for these errors on the part of the subjects.

Table 1

Results of Series I

Data	Calls	Dev.	CR	P
First Half	3142	+64.6	2.88	.004
Second Half	3136	− 1.2	.05	.96
Total	6278	+63.4	2.00	.045

The machine scoring of Series I indicated that the 21 subjects had scored a total of 1,319 hits. The actual number of calls was found to be 6,278 (see Table 1). The expected number of hits is one-fifth this number, or 1,255.6, and the observed total score thus shows a deviation of 63.4 over "chance." The deviation on the first half of the series (the first page of 150 calls by each subject) was +64.6 for 3,142 trials, while that for the second half was −1.2 for 3,136 trials. Thus the first half contributes all of the positive scoring.

In Series II each of 63 subjects was supposed to make 150 calls. There were found 49 sets that were unmarked and two with two choices, so that the total number of trials was 9,399 (see Table 2). These gave a total score of 1,955 hits, or a deviation from "chance" of +75.2. The first half of the series gave a deviation of +20.4 on 4,698 trials and the second half a deviation of +54.8 on 4,701 trials. The deviation was contributed about equally by those who answered yes and those who answered no on the task set question. The comparison of the results for the first and second halves of the second series showed a slight but insignificant increase in the scor-

Table 2

RESULTS OF SERIES II

Data	Task Set	Calls	Dev.	CR	P
First Half.....	Yes	3209	+ 5.2		
	No	1489	+15.2		
	Total	4698	+20.4	.74	.46
Second Half.....	Yes	3207	+29.6		
	No	1494	+25.2		
	Total	4701	+54.8	2.00	.045
Entire Series....	Yes	6416	+34.8		
	No	2983	+40.4		
	Total	9399	+75.2	1.94	.052

ing rate. The contribution to the total results by the subjects who did not expect to score above "chance" was about twice as great as that of those who did, but again the difference does not approach statistical significance.

The usual method of evaluating ESP data is based upon the binomial formula. By this method Series I gave a CR of 2.00 (P = .045); Series II gave a CR of 1.94 (P = .052); and the combined results gave a CR of 2.77 (P = .006).

Since the results of these two series are based on tests in which many subjects made calls for the same targets, it must be recognized that the binomial method of evaluating the data may not be appropriate. If there were any group tendency among the subjects that led them to mark the same choices on the same trials, the binomial formula would overestimate the significance of the results. A method for evaluating multiple calling data of this sort without making any assumption regarding independence among the responses has been presented by Greville (1).

In order to have the results re-evaluated by this method, the data were sent to the Parapsychology Laboratory of Duke University, where they were completely checked and rechecked under the supervision of Dr. J. G. Pratt. (The total score obtained by the hand tabulation agreed with that given by the machine.) The appropriate analysis was then carried out under the direction of

Dr. T. N. E. Greville. The values of the CR's obtained were slightly different from those obtained on the assumption of independence (see Table 3), but the results of both series combined are still significant (CR = 2.51; P = .012).

The reader will note that in Series I both the total deviation and the difference between the halves were in the expected direction though neither was significantly so. In Series II, the total deviation was also positive. However, no decline effect was found, nor did the verbal responses to the preliminary question correlate significantly with the results. The results of the two series, combined as shown in Table 3, yield a deviation from the expected value which is not attributable to chance.

Table 3

COMBINED RESULTS

Data	Calls	Dev.	BINOMINAL		GREVILLE METHOD	
			CR	P	CR	P
Series I.....	6,278	+ 63.4	2.00	.045	1.85	.064
Series II.....	9,399	+ 75.2	1.94	.052	1.74	.082
Total.....	15,677	+138.6	2.77	.006	2.51	.012

CONCLUSIONS

Evidence of ESP has been obtained with a procedure adapted to machine scoring. The present application of the method, however, reveals a weakness in the procedure which was not detected by the mechanical method. This is the possibility of mistakes by the subjects in marking their calls. These mistakes were discovered only in the recheck at Duke University. Sixty errors of omission (not marking a given set) were detected, and 13 errors of commission (making two marks in a given set) were found. Because of the predominance of errors of omission over errors of commission in the series reported here, the results obtained by the machine check were lower than those obtained by hand tabulation. Errors of omission, if undetected, lower the deviation since they increase the apparent number of calls without changing the number of hits. Errors of commission, however, tend to increase the apparent significance of the results since more than one mark in a set increases the pos-

sibility of a hit without a proportionate increase in the apparent number of trials. Subjects' errors in these particular series led to an underestimation of the significance of the results by the machine-scoring method.

In future series it cannot be assumed, of course, that errors of omission will outweigh those of commission. Unless some method is devised whereby these errors can be detected by the machine, it will be necessary for the experimenter to count them. It is probable that more explicit instructions to the subjects would reduce the extent of the errors and thereby minimize the possibility of personal mistakes on the part of the experimenter. The machine method does provide a great saving in time since a large number of trials can be scored for hits very quickly. The method would appear to be especially advantageous for rapid exploratory work where immediate knowledge of results is desired. In view of the results of the present experiment, it seems unlikely that highly significant results obtained by machine scoring would be reduced to insignificance by a later study of subjects' errors.

The experimental differences between the two series merit some consideration. (1) The subjects in the first series were selected on the basis of the above-"chance" performance in previous testing without regard to their attitude (sheep or goat). (2) On the other hand, subjects in the second series were selected only insofar as they were all sheep. (3) In the first series, the subject-target distance was approximately one mile; in the second, it was 500 miles. (4) There was a considerable age difference between the groups.

Since the subjects in the second series were all sheep, it is interesting to compare the results of this series with the over-all averages last reported by Schmeidler (2) for sheep. The 9,399 calls in the second series are equivalent to 375.96 runs with ESP cards. The 1,955 hits thus yield an average of 5.20 hits per run, which compares favorably with Schmeidler's figure of 5.15.

The basic hypothesis involved in the problem of task set is that the goal which the subject sets for himself on an ESP test will correlate positively with his actual score. We attempted to test this hypothesis, making the assumption that if a subject were compelled to make a verbal statement of his own expectation of scoring above or at "chance," he would as a rule tend, consciously or uncon-

sciously, to set his actual goal in accordance with his statement. The results do not bear out the hypothesis. Accordingly we must conclude either that the basic hypothesis is in error or that the particular verbal response which we elicited did not in reality reveal or establish the task set of the individual. Naturally, these conclusions must be considered as only tentative in view of the relatively small number of subjects tested.

Experiments involving the IBM machine seem to differ qualitatively from most other GESP experiments in at least one important respect. The task facing our subjects was to intuit the *spatial location* of marks on a page; that is, to intuit in which of five rectangles a black mark was located. There are two distinct ways in which the subject might accomplish this. He might intuit each set (out of the 300) as a separate unit, or he might intuit configurations of marks, or patterns across the page. Considering the nature of the experiment we have assumed that some form of clairvoyance is a more probable explanation than an unconscious telepathic effect on the part of the experimenter who made up the target sheet. This place-perception situation is in contrast to that in most GESP tests in which the subject is asked to make successive distinctions between different kinds of geometric symbols. Here only one symbol, a black mark almost completely devoid of affective connotation, is used. Its spatial orientation is a crucial factor. At this point, of course, it is impossible to tell whether these target distinctions are related to the paranormal capacity of the percipient. This is a problem which should be put to experimental test.

References

1. Greville, T. N. E. On multiple matching with one variable deck. *Ann. math. Statist.*, 1944, **15,** 432-34.
2. Schmeidler, G. R. Personality correlates of ESP as shown by Rorschach studies. *J. Parapsychol.*, 1949, **13,** 23-31.

Harvard Society for Parapsychology
Harvard University
Cambridge, Massachusetts

MINOR ARTICLES

REPORT ON AN EXPERIMENT CONCERNING THE INFLUENCE OF MIND OVER MATTER

By E. A. G. KNOWLES

ABSTRACT: This PK experiment with dice was first started as a means of providing illustrative data for a lecture on applied statistics. The subjects were the author, several colleagues, and nine members of the lecture audience. Twelve dice were thrown at a time by each subject. Approximately half of the time the subject threw for the high faces (4, 5, and 6) and half for the low faces (1, 2, and 3). The experiment as a whole gave results significantly better than chance. Later a loaded die was used in a control series in which the subjects did not try to use PK. The results from the loaded die tests were indistinguishable from those with the normal dice in which the throwers did try to use PK.

The author, Dr. E. A. G. Knowles, is an engineering mathematician employed by a firm of English engineers.—Ed.

CONDITIONS AND BACKGROUND OF THE EXPERIMENT

To begin with I should point out that I am not engaged in psychical research, nor am I a member of any society carrying out such research. I am an engineering mathematician, interested in the application of probability theory and statistics in processes of mass production.

I had, rather incredulously, heard from my brother of successful experiments in the U.S.A. on the effect of concentration on the scores achieved by tossing dice. However, I have not read any reports or literature on the subject. I took part in what I considered an unsuccessful experiment of my brother's in tossing twelve dice, which convinced me that the dice were reasonably unbiased, so that I could safely use them in experiments to illustrate sampling theory.

In particular I was preparing a lecture and discussion entitled "Probability, Sampling, and Significance" for members of my firm, mainly for engineers. While doing this, it occurred to me that such a lecture could be made much more interesting if everyone took part in an experiment of which no one could know the answer beforehand. I proposed to illustrate a significance test on material to which each one had contributed.

Thus, I saw my assistant, Miss D.G., who was preparing other material for the lecture, told her of my purpose and of the fact that successful experiments had been conducted at an American university on the influence of concentration on the fall of dice. However, I said, how could we believe this if we had not seen it happen or proved the results significant? I said that there were many who did not believe the results and that the answer was probably not yet fully known and required more experiment. She was very keen to try it, and she and her four colleagues in the measurements laboratory soon got together to try out the first set of trials. I had explained that I wanted a very quick way of recording the results so that, with their help at the lecture, we could get as many data as possible while we were waiting for people to assemble. I said that I wanted each throw of the dice recorded separately in the order of occurrence and, so that no information should be lost, a record of how many of each face turned up each time, as well as who was throwing the dice. As in my brother's experiment, twelve dice were thrown at one time from a bakelite beaker, each person shaking the beaker well before throwing the dice. They were thrown onto an ordinary wooden table during the course of the recorded experiments. The results found in the two sets of trials, (1) the preliminary tests done by my colleagues before the lecture and (2) the data collected at and after the lecture, are given in Table 1 and Table 2, respectively.

Preliminary Trials

Five persons were present at the preliminary trials, including Miss D.G. They decided that each one should throw the dice once, trying for low numbers, assisted by the mental efforts of the rest. When all had had one turn, all then tried for high numbers, while each in turn threw the dice.

During this initial period of throwing with high aim, one person was called away, so that only four throws were recorded. After this, all four had another turn at trying for low numbers; but before they could carry out the next set of throws with high aim, they were interrupted. Thus the first set of trials was ended, making, in all, thirteen trials on twelve dice per trial, nine aimed low and four aimed high. The following table shows the results in brief. The

Table 1

RESULTS OF THE PRELIMINARY SERIES

	Number Lower Faces (1, 2, 3)	Number Higher Faces (4, 5, 6)	Total
Low-Aim Throws	67 (60.9)	41 (47.1)	108
High-Aim Throws	21 (27.1)	27 (20.9)	48
Total	88	68	156

Approximately $\chi^2 = 3.81$ and $P = 0.051$

numbers in parentheses in this and the following tables show the average expected frequency in each cell of the fourfold table.

The shaded parts of the histograms given in Figure 1 show the results obtained for each face. The reason for dividing the throws into only two categories according to whether they showed face 3 and below or face 4 and above was the fact that all felt they were having success if, on aiming high, they had results above 3 and on aiming low, results below 4.

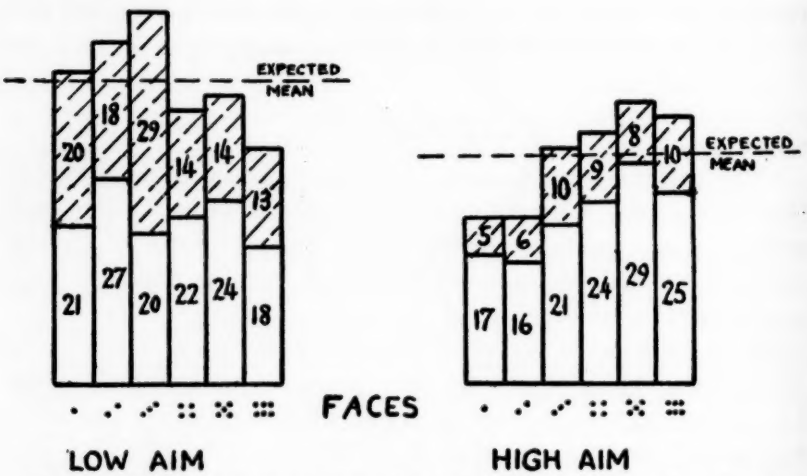

FIG. 1. Histograms showing results for each face of the die. The preliminary tests are represented by the shaded areas; the tests performed at the lecture, by the unshaded areas. The numbers show the actual frequencies for each face under each condition.

When using the tables by Fisher and Yates (1) it turned out that the value obtained for P was just above the 5% level of significance. So the results could not be claimed significant. However, in normal engineering experimental trials, this result would be considered encouraging and further evidence would be sought with the hope of increasing the significance.

We were all well pleased with such a lucky result, and I therefore thought that, as no one would believe in the possibility of influencing the fall of dice by one's mental efforts, the example might well serve as a warning against jumping to conclusions on the basis of insufficient evidence! However, I wanted to verify first that this condition of disbelief was really fulfilled. For this reason the next set of trials was carried out in the way described below.

Trials At and After the Lecture

This set of trials was organized by the group which carried out the first set of trials and was watched from all sides by interested onlookers. On entering the lecture room each person was asked to answer the following questions:

1. Do you believe that it is possible to influence the fall of unbiased dice by the application of one's will or wish?

2. Do you believe that it is *im*possible to influence the fall of unbiased dice by the application of one's will or wish?

3. Do you believe that this problem is not a matter of belief, but should be answered by experiment?

Out of 21 persons who answered the questions, no one answered yes to the first; 12 answered yes to the second; and 13 answered yes to the third.

The completion of this form took longer than I had expected for the reason that the third question appealed rather strongly to the scientific conscience and was thus a cause for thought and hesitation. Unfortunately, therefore, not all present were able to try the dice experiment. Only nine people took part in this second set of trials. No one made any trials before having completed the questionnaire. The subjects were instructed to aim high for the first throw and aim low for the second throw. One of them made three trials each of aiming high and aiming low, so that in all there were 22 trials with 12 dice per trial, 11 with high aim and 11 with low aim. The

results are given in Table 2, while the unshaded portions of the histograms in Figure 1 give the details for each face. The number of successes was much lower than that for the first set of trials. Thus we thought that our experiment had led to insignificant results.

Table 2

RESULTS OF TRIALS MADE AT THE LECTURE

	Number Lower Faces (1, 2, 3)	Number Higher Faces (4, 5, 6)	Total
Low-Aim Throws	68 (61)	64 (71)	132
High-Aim Throws	54 (61)	78 (71)	132
Total	122	142	264

Approximately $\chi^2=2.58$ and $P=0.11$

SUMMARY OF RESULTS

Several weeks later, one evening, I had another surprise. I thought that just for the sake of completeness I might as well add up the results of the two sets of trials. Thus I obtained the figures shown in Table 3. The results of the summation are also shown in the histograms of Figure 1, taking the shaded and unshaded portions together.

Table 3

COMBINED RESULTS OF PRELIMINARY TRIALS AND THOSE DONE AT THE LECTURE

	Number Lower Faces (1, 2, 3)	Number Higher Faces (4, 5, 6)	Total
Low-Aim Throws	135 (120)	105 (120)	240
High-Aim Throws	75 (90)	105 (90)	180
Total	210	210	420

$\chi^2=8.18$
$P=0.0042$

The statistical significance test led to a value of P which was less than .005, a result generally considered very significant in engineering trials.

It may be of interest to note that in the preliminary set of tests the low-aim trials were done first by each thrower, and in the second set, the high-aim trials were done first. It happened that in the preliminary set of trials there was relatively more success at aiming low, and in the second set, more at aiming high. The significance of the combined results was largely due, therefore, to the throws for the faces first aimed at in each set. In this outstanding success on the first trials of the subjects in each set, the results are, I understand, like those obtained in a number of PK experiments previously reported.

In case the significance of the above result is due to a psychological effect and not to an unusual chance occurrence, it would be of interest to note the differences in psychological conditions in my own experiment as compared with the PK tests carried out by my brother. In my own experiment the spirit of play and of "trying anything once" was alive in the group, and as the time of experiment for each individual was very short, no one had to concentrate for sufficiently long to become tired of the effort. My experiment was a group experiment, designed to get as much as possible out of the least individual effort. My brother's experiment was carefully designed to test the significance of the individual effort. For this reason each individual had to concentrate for at least one hour in order to produce the desired sample size. Therefore, although the individuals in my brother's group initially had a similar playful spirit in some cases a really great interest and enthusiasm for the subject itself, there can be no doubt that the difference in psychological conditions for a prolonged test as compared with a short test are very great and might account for the difference in results.

Comparison with Experiment Using a Loaded Die

In view of the results described, it appeared of interest to obtain a similar set of results that were due entirely to physical factors in the situation. This would permit some sort of estimate of the physical forces that would be required to yield results similar to those of the experimental series.

Since "low aim" in the previous experiment meant wishing for numbers 1, 2, and 3, a die was loaded at the corner which the faces 4, 5, and 6 have in common. Thus, on throwing the die, a bias in favor of the appearance of 1, 2, or 3 as the uppermost face was to be expected.

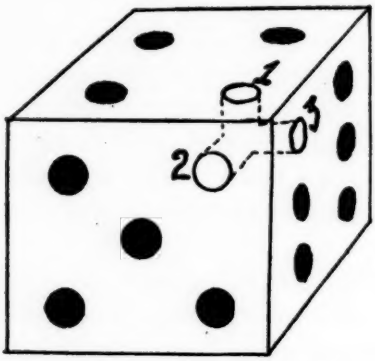

FIG. 2. Diagram showing manner of loading die.

Details of Die and Load Used

On the 4-, 5-, and 6-faces three holes (marked 1, 2, and 3 in Fig. 2) were drilled through the three dots nearest the corner that was common to the three faces. The holes were the same diameter as the dots and the depth was just great enough for all three holes to meet. The hollow thus formed was filled with lead. The end faces of the lead were painted black in order to make them appear like the others.

The relevant characteristics of the die before and after loading are as follows: The original weight of the die was 3.26 grams. After loading it was 3.62 grams, the increase in weight being approximately 11%. The distance between parallel faces was 14 mm. The diameter of the dots was 2¼ mm. The distance of the center of the dots from the edges was 3½ mm.

Conduct of the Experiment

In all, eight people took part in producing results for the frequencies of occurrence of the different faces. The die was thrown for 905 times from a cup. All except one person (myself) contributed 100 throws each. All persons, other than myself, were ignorant of what had been done to the die, and, since some of them

had taken part in the previous experiment, they were asked not to try to influence the dice. Two of them were quite relieved not to have to do so, since they had tried dice-throwing tests of PK independently without success. Since the results were produced by different people at different times and in different places and since no one had any conscious wishes, it seems unlikely that any consistent psychological effect should have existed to influence the results, assuming that such influence is possible. In order to check on this, each person carried out 100 throws with a die of the same sort, but unloaded, before using the loaded die.

Results of Control Experiment

As I had hoped when preparing the die, the throws made with this loaded die gave results that consistently favored the low faces. Furthermore, the amount of favoring with this special die appeared to be in about the same proportion as that obtained in the PK throws. The detailed results of the control series with both the

Table 4

RESULTS FROM THROWING A LOADED DIE AND A NORMAL DIE WITHOUT ATTEMPTING TO USE PK

Throwers	Number Lower Faces	Subtotals	Number Higher Faces	Subtotals	Totals	Subtotals
LOADED DIE						
D. G.	55		45		100	
J. A.	61		39		100	
F. K.	25	141	15	99	40	240
F. K.	38		22		60	
J. S.	56		44		100	
E. K. 1	8	102	12	78	20	180
E. K. 1	50		30		80	
E. B.	61		39		100	
M. W.	66		34		100	
F. G.	65		40		105	
E. K. 2	60		40		100	
Grand Total Loaded Die	545		360		905	
UNLOADED DIE						
Results for Unloaded Die	408		392		800	

normal die and the loaded one are shown in Table 4. It will be observed that the result for the unloaded die is consistent with chance expectation for an unbiased die.

It was desirable to prevent the necessity of making another set of experiments with another die so loaded as to give a bias in favor of 4, 5, and 6 as this could not easily be done without the loading being very noticeable because of the location of the dots. Hence the results of Table 4 were used in the following manner in order to obtain from the control series samples of both high- and low-aim throws of the same size as the PK series shown in Table 3.

Two complete sets of 100 throws were taken (D.G. and J.A.), and the next 40 from the set of the next person (F.K.) were combined with them to be compared with the 240 low-aim results of Table 3. Then I assumed that for the next 180 throws shown in the table the faces of the die had been renamed so that every time the 1-face appeared it was recorded as a 6, the 2-face as a 5, etc. In this way we were able to get 180 throws with the loaded die for high aim by taking the remainder of 60 (F.K.) and combining these with the next set of 100 (J.S.) and with 20 from the next on the list (E.K.1). These 180 scores were thus written down in reverse order to that shown in Table 4; that is, the 102 throws that gave faces 1, 2, and 3 were put under 6, 5, and 4, and conversely. The figures from the loaded die experiment used in this comparison are those shown in the Subtotal columns of Table 4. Thus Table 5 was obtained for comparison with Table 3.

Table 5

Results of 420 Biased Die Throws

	Number Lower Faces (1, 2, 3)	Number Higher Faces (4, 5, 6)	Total
Low-Aim Throws	141 (125.1)	99 (114.9)	240
High-Aim Throws	78 (93.9)	102 (86.1)	180
Total	219	201	420

$\chi^2 = 9.19$
$P = 0.0024$

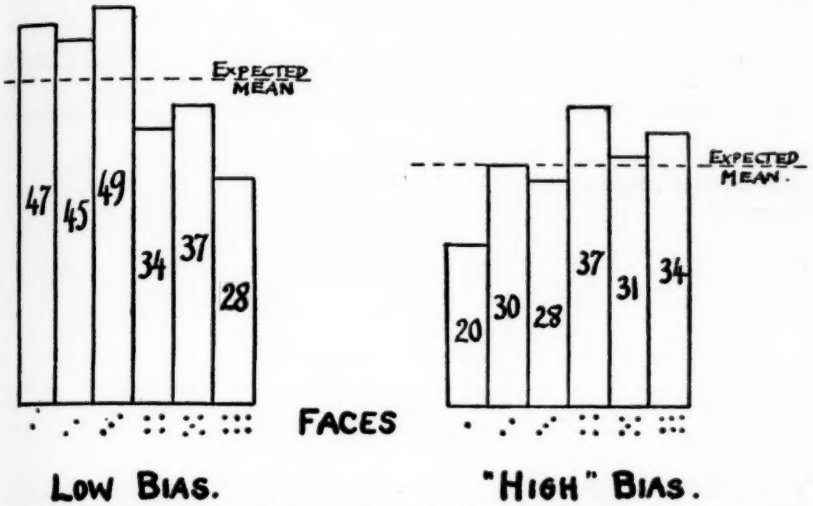

FIG. 3. Frequency of faces for tests with a die loaded to favor low faces. The distribution of 240 throws with this die is given on the left; the distribution of 180 additional throws (arbitrarily selected to represent a die loaded for "high" bias) is shown on the right.

It must be admitted that these results are surprisingly similar to those of Table 3. Figure 3 represents the frequency diagram for the individual faces for comparison with Figure 1.

In order to get a quantitative comparison of the experimental and control results, I used two statistical measures. First, using the coefficient $V = \sqrt{\chi^2/N}$ as the measure of association (2, pp. 309, 311) in order to compare the results of Table 3 with those of Table 5, it is found that approximately

$$V = 0.140 \text{ for Table 3}$$

and

$$V = 0.148 \text{ for Table 5}$$

The two results are thus virtually identical.

So far the complete sample for the loaded die (the "Grand Total Loaded Die" of Table 4) has not been used. The comparison of the data from the loaded die with the results of the experimental series is shown in Table 6.

The 240 successes in the experimental series were obtained from Table 3 by adding together the 135 low-aim results for the 1-, 2-, 3-faces and the 105 high-aim results obtained when throwing for the 4-, 5-, 6-faces. Similarly, 180 was obtained by adding the corresponding numbers for unsuccessful throws from the same tables.

Table 6

COMPARISON OF TOTAL RESULTS WITH LOADED DIE WITH COMBINED EXPERIMENTAL RESULTS OF PRELIMINARY AND LECTURE SERIES

	Exper. Series (Unbiased Dice)	Control Series (Loaded Die)	Total
Successes	240 (248.8)	545 (536.2)	785
Failures	180 (171.2)	360 (368.8)	540
Total	420	905	1325

$\chi^2 = 1.26$
$P = 0.26$

The fact that these numbers are identical with two marginal totals of Table 3 is merely an accident.

Using the same method of evaluation as before, it was found that $P > 0.2$. Hence, statistically, there is no evidence that the two samples have not been derived from the same population.

CONCLUSION

Although the latter experiment with the loaded die cannot, of course, prove anything about the nature of the first results obtained while the subjects were trying to influence the dice, it does give a rough idea of the magnitude of a set of physical forces that might be involved on the assumption that the observed distribution is *not* due to chance but is typical of the psychological process in question.

REFERENCES

1. FISHER, R. A., AND YATES, F. *Statistical Tables for Biological, Agricultural, and Medical Research.* Third Ed. London: Oliver and Boyd, 1948.
2. KENDALL, M. G. *The Advanced Theory of Statistics.* Vol. 1. London: 1945.

Orchard Street
Abergavenny
Monmouthshire, England

A CLAIRVOYANCE GAME EXPERIMENT

By WILLIAM R. BIRGE

ABSTRACT: This clairvoyance experiment was planned to encourage an attitude of spontaneity among the subjects in the hope that this would give more successful ESP results. The tests were designed as six different card games which the subjects played together at each session. Small prizes were given for winning scores. The results gave only suggestive evidence of ESP.—Ed.

INTRODUCTION

FOR a long time students of parapsychology have felt confident that the following characteristics may safely be said to be true of ESP: (1) that test scores are likely to decline within the run, session, or series; (2) that a gamelike atmosphere of spontaneity is conducive to high scoring in ESP tests. The present experiment was designed to secure maximum positive scoring by capitalizing on these well-established features of the extrasensory process. In order to promote spontaneity, a procedure was adopted which involved the embedding of the ESP task in an actual game situation, with small prizes for good scores; and to inhibit decline effects the structural unity of each run of twenty-five calls was broken up into five clearly demarcated segments containing five calls each. Although the experiment yielded only suggestive results, it is presented here largely because of the new technique of ESP testing which was employed.

PROCEDURE

The research was carried out at the Duke Parapsychology Laboratory from October, 1946, to May, 1947. The experiments consisted entirely of tests of clairvoyance, and the standard ESP cards were the basis of all the tests. Six subjects participated in each experimental session, three subjects being seated on either side of a long rectangular table. Each subject was provided with a pencil and a small lined grocery pad containing carbons on the back of every other page. Stripped of its game variations, the task of each subject was simply to call down through twenty small packs of five ESP cards arranged face up and enclosed in opaque envelopes.

More specifically, the experimenter (W.R.B.) would bring with

him to each experimental session twenty opaque envelopes, each containing five ESP cards. These envelopes had been prepared in advance by an assistant who had selected the enclosed target cards on a random basis; that is, a dice-throw-determined entry into Tippett's tables of random numbers. A code devised by the assistant was used to change the numbers in Tippett's tables to ESP symbols. Through these precautions a random order of target cards was insured. At the same time, having the assistant place the cards in the opaque envelopes made it impossible for the experimenter to give the subjects sensory cues of any kind. The flap of each envelope was held down by means of a paper clip.

The subjects were told that a particular envelope might contain any possible combination of the five symbols, ranging from five cards all with the same symbol to five cards of different symbols.

A uniform procedure of presenting the target cards and recording the calls was maintained throughout each experimental session. The experimenter would enter the room, seat himself at the head of the table, and place the twenty opaque envelopes on a small table directly in back of him. He would then give the necessary instructions regarding the ESP symbols, games, recording, prizes, etc. Next he would reach back, select any one of the twenty envelopes, and place it on a sheet of white paper in the center of the table so that the cards within were face up. The subjects then recorded their impressions of the order of the five cards on the left-hand side of the first sheet on their pads. They tore off these sheets after their five calls had been recorded and passed them to the experimenter, who clipped the six records together and placed them in a locked box at his side. Then the experimenter opened the envelope and read off the correct order of the cards. Each subject recorded the true card order on the right-hand side of the second sheet of his pad. Since every odd-numbered sheet on the pads had a carbon back, the subject's calls appeared on the left-hand side of the second sheet, and the subjects were able to make an immediate check on hits and misses. After the scoring had been completed and the "banking" called for by the rules of the game had been completed, the experimenter would place another envelope in the center of the table. The same procedure as outlined above would then be repeated. The subjects would turn over the second page of their record pads and

record their calls for the second envelope on page 3 of these pads. By the end of the twentieth envelope each subject had called the equivalent of four runs of twenty-five calls. Thus twenty-four runs of data were collected from the group of six subjects at every experimental session. All record pads were, of course, turned in to the experimenter at the end of each session.

It was clearly impossible for a subject to falsify his score, since each subject was required to surrender his call sheet before the target envelope was opened; and of course it was against the original call sheets that the target card orders were later independently checked. Furthermore, multiple recording by all six subjects of the target card orders insured the accurate recording of the target cards. Wherever discrepancies among target card recordings existed, the judgment of the majority was assumed to be correct. Since in no instance did the records show more than two discrepancies in the recording of any one target card, the method appeared to be empirically justified.

Games

Four kinds of games were played during each experimental session, each game involving the calling of five envelopes of cards. Common features of all of the games were the winning of poker chips by successful subjects in the course of a game, and the awarding of a small prize, usually a candy bar, to the subject or subjects who had received the most poker chips at the conclusion of each game.

There were actually six games employed in the course of the entire experiment. A game entitled "Extra Squared" was played only during the first two sessions of Series I, while one game, "Old Metapoker," was played only during the first session of Series I. These two games were replaced with what appeared to be better games. The four games used extensively in the experiment were named "Superodds," "New Metapoker," "Parapoker," and "Scoop."[1]

The entire experiment consisted of three experimental series. The conditions wherein the three series differed will be given briefly in the next section along with separate reports of the results for each series.

[1]Since the results of the experiment were only suggestive, it is not considered worth while to present full descriptions of each of the games. Anyone planning to conduct a similar research project may obtain the details by writing to the Parapsychology Laboratory.

Results

Series I

Series I was an exploratory series. The experimenter therefore allowed himself complete freedom to: (1) replace old games with more promising ones; (2) alter the order of game presentations; and (3) select either six-men, six-women, or three-men-three-women combinations of subjects. It was believed that only through a flexible, trial-and-error procedure could the optimum conditions for positive scoring be found. Also, it was considered more profitable to seek questions rather than answers from the data of the first series.

There were ten experimental sessions in Series I. Games, game-order presentations, and subjects were selected in the following manner: Superodds and Parapoker were played during every session. Extra Squared was succeeded by Scoop after the first session, while Old Metapoker was dropped after two sessions and was replaced by New Metapoker. The order of game presentations was varied unsystematically until the seventh session. From the seventh through the tenth sessions, the order of game presentations was Superodds, Parapoker, New Metapoker, Scoop.

Subjects were selected almost exclusively from the undergraduate population of the university (as was also the case in the next two series). All of the subjects participating in the first six sessions were girls. Only men subjects were used in the seventh and eighth sessions, while the ninth and tenth were "mixed" sessions. At "mixed" sessions, three of the subjects were women and three were men.

At the conclusion of Series I, the data were first analyzed in terms of direct hits on target cards, and forward and backward displacement scores. A subject's call which matched a just-succeeding or a just-previous target card was considered a forward or a backward displacement hit respectively. When the results were analyzed in terms of runs of 25 cards each, the following figures were obtained:

	Number of Runs	Deviation
Direct	240	+11
For. Displacement	192	+21
Back. Displacement	192	−48

These deviation scores do not differ significantly from chance expectation. Neither did the direct scores show significant differences when they were broken down into the different games. With the omission of Extra Squared and Old Metapoker, games which were played only once and twice respectively, the results by games were as follows:

	Number of Runs	Deviation
Superodds	60	−14
Parapoker	60	− 4
New Metapoker	48	+15
Scoop	54	+24

A more encouraging feature of the results was an analysis of the direct hits in terms of the serial order of the four runs within a session. The following figures show the incline in scoring which was obtained when all the hits for each of the four run positions within the session were summed and pooled:

Run Position	Number of Runs	Deviation
1	60	−22
2	60	+ 1
3	60	− 1
4	60	+33

The critical ratio of the difference between the first and fourth run position scores is 2.51. Since this CR is based on a feature of the data selected post facto, it cannot be considered significant. However, it seemed sufficiently impressive to justify the planning of the next series with the view to answering the questions which this result raised.

Series II

With reference to the results of Series I, the hypotheses which might account for the observed incline in scoring rate during the session were considered to be three in number: (1) chance fluctuations; (2) the building-up of rapport during the session; (3) superiority of the game Scoop, which was played in the last run position during four of the sessions and which accounted in large part for the positive deviation obtained on the fourth run. Series II was designed to test experimentally which of these three hypotheses was most likely to have been correct.

Series II consisted of twelve experimental sessions of mixed (that is, three men-three women) groups. The sessions were arranged in such a way that each of the four games, Scoop, Superodds, Para-

poker, and New Metapoker, was played for three sessions in each of the four run positions. This was accomplished by systematically advancing each game presentation one run position in successive sessions. For example, the order of game presentations in Session I was Scoop, Parapoker, New Metapoker, Superodds. In accordance with the rotation scheme outlined above, the order of game presentations in Session II was Superodds, Scoop, Parapoker, New Metapoker. Thus at the end of the twelve sessions, each game had been presented three times in each of the four run positions.

Coming back to our three hypotheses, it was believed that the chance hypothesis would be ruled out if significant results supporting either (or both) of the other two hypotheses should be obtained. If the last run should show a consistent superiority regardless of which game was played in that run position, the "increased rapport" hypothesis would definitely be indicated. Finally, should the game Scoop yield consistently high scores in each of the four run positions, it would seem reasonable to assume that the superiority of Scoop accounted in some measure for the incline effects found in Series I.

The analysis of the direct and displacement scores of Series II is summarized below:

	Number of Runs	Deviation
Direct	288	+47
For. Displacement	230.4	−45
Back. Displacement	230.4	+69

While none of the three deviation scores is individually significant, the critical ratio of the difference between the forward and backward displacement scores is 2.6. Here again the CR cannot be regarded as significant since the evaluation was based on an aspect of the data selected after the data were examined.

When the direct hits were analyzed in terms of games, the results of Series II were consistent with those of Series I. The results of this analysis were as follows:

	Number of Runs	Deviation
Superodds	72	− 2
Parapoker	72	− 1
New Metapoker	72	+32
Scoop	72	+18

As in Series I, Scoop and New Metapoker appear to have been

distinctly superior to Superodds and Parapoker. When the scores of the two series are combined, the CR of the difference between the two pairs of games is 2.6. Since such an evaluation is unduly selective, however, the difference can only be regarded as suggestive.

Analysis of the game Scoop revealed a small but consistent positive deviation for each run position. With each deviation score being based on 18 runs, the Scoop results by run position were: +1, +9, +1, +7. Altogether, the results suggest that chronological position within the run is probably of less consequence than the differences between the individual games, and that chance (that is, indeterminate) factors are of signal importance in a game situation.

There were other respects in which the results of Series II showed no regard for consistency. In contrast to the position effect pattern of Series I, the scores by run position within the session were as follows:

Run Position	Number of Runs	Deviation
1	72	+20
2	72	+ 2
3	72	+17
4	72	+ 8

Series III

In Series III, as in Series II, the order of presentation of the games was systematically varied and each one was played three times in each of the four run positions. In Series II the two most successful games, Scoop and New Metapoker, were never played in succession but were separated by an intervening game. In order to obviate the effects this separation may have had on the superiority of these two games, Scoop and New Metapoker were presented in successive run positions in Series III. The order of presentation of the games in the first session of the series was as follows: Parapoker, Scoop, New Metapoker, Superodds.

Series III was comprised of twelve experimental sessions. Six sessions were made up of mixed groups, while the remaining six unmixed sessions included five all-men and one all-women sessions. Mixed and unmixed sessions were alternated throughout the series. The difference in the number of all-men and all-women sessions was necessary in order that the total experiment (including all three series) might have an equal number of all-women and all-men sessions.

The direct and displacement deviation scores obtained in Series III are given below:

	Number of Runs	Deviation
Direct	288	−14
For. Displacement	230.4	−56
Back. Displacement	230.4	+25

All of these scores are well within the range of chance fluctuations, and perhaps their only interesting feature is the sizable difference between the displacement scores. A similar and even more striking difference, it will be recalled, was found in Series II.

When broken down by games, the direct deviation scores were as follows:

	Number of Runs	Deviation
Superodds	72	+ 5
Parapoker	72	+ 6
Scoop	72	− 8
New Metapoker	72	−17

Of particular interest in the above data is the fact that both Scoop and New Metapoker, previously outstanding games, finished behind the other two games in direct scoring.

Nothing of significance was revealed by the analysis of run scores in terms of run position or serial order within the session. The results of this analysis are as follows:

Run Position	Number of Runs	Deviation
1	72	−24
2	72	+ 2
3	72	+16
4	72	− 8

These position effects are not significant, nor are they similar to those found in the two earlier series.

Summary of Results

The results of the three series combined are given in Table 1. The lack of significance in the results can be determined by inspection. In addition to the studies presented in the table, however, a test of significance was made on the differences between the forward and backward displacement scores of the three series. The results of this study are presented below:

	Runs	Dev. (For. Displace.)	Dev. (Back. Displace.)	Diff.	SD_d	CR_d	P
Series I	192	+21	−48	69	39.19	1.76	.0784
Series II....	230.4	−45	+69	114	42.93	2.65	.0080
Series III...	230.4	−56	+25	81	42.93	1.89	.0588

Combined by Fisher's method the probabilities shown in the table for the three series give a P-value of .002 for the experiment as a whole.

The wide divergence of the displacement scores was an unexpected occurrence, and of course the testing of this divergence represented only one of several analyses. Also, Fisher's method is based on the assumption of the independence of the forward and backward displacement scores. Since there is a slight interdependence between these two scoring categories, this assumption is not fully justified. Thus the low probability figure indicates a trend of scoring pattern which could only be regarded as significant if it were to hold up in further experimentation.

Table 1

COMBINED RESULTS OF ALL THREE SERIES

Target Card	Number of Runs	Deviation
Direct	816	+44
For. Displacement	652.8	−80
Back. Displacement	652.8	+46
Games		
Superodds	204	−11
Parapoker	204	+ 1
Scoop	198	+34
New Metapoker	192	+30
Extra Squared	12	0
Old Metapoker	6	−10
Run Position		
Run 1	204	−26
Run 2	204	+ 5
Run 3	204	+32
Run 4	204	+33
Subject Combinations		
Mixed	480	+56
All-Women	168	−21
All-Men	168	+ 9

DISCUSSION

The experiment under discussion was designed to secure the optimum test conditions for the operation of clairvoyance. For the purpose of providing an atmosphere of maximum spontaneity, each test run was embedded in a game context. The target cards of each run were distributed among five separate envelopes with a view to minimizing decline effects which might result from the run structure. In spite of the care with which the stage was set, the results

show little evidence of the occurrence of ESP. Of the various studies made on the data, only the chi square analysis of the difference between the forward and backward displacement scores turned out to be significant. Since this aspect of the data was both selected and unexpected, it does not suffice to sustain a conclusion that ESP occurred in this experiment.

On the basis of the research literature, it is difficult to account for the present negative results in terms of deficiencies in the objective experimental conditions. It seems more likely that the fault resided either within the subjects, who may have been endowed with very limited ESP capacity, or with the experimenter, who may not have provided the proper inspiration or other psychological atmosphere which is conducive to the functioning of ESP.

The only suggestive evidence of ESP in this study was provided by the displacement scores. In a recent report (1) the writer indicated that an element of frustration was found to exist in many of the displacement experiments. Although it is little more than a speculative hunch, the experimenter believes that his own excessive eagerness may have inhibited the ESP of his subjects. For example, in Series II he keenly anticipated finding the same run position effects which he found in Series I. In Series III he had great hopes that the games Scoop and New Metapoker would continue to elicit positive scoring. Neither of these expectations was fulfilled, and it is at least plausible to assume that the tensions aroused by these anticipations may have created a psychological situation which insured their failure.

It will be recalled that the experimenter took only an administrative and supervisory part in the playing of the games. If the function of the ESP games was to dispel the tensions and self-consciousness of the participants, perhaps the negative results in this experiment were caused by the failure to provide a method whereby the tensions and self-consciousness of the experimenter might be reduced. This reasoning suggests that in research with ESP games, the experimenter himself would do well to join in the games as an active participant. Certainly one of the major contributions of the experiment is the sharp manner in which it points up

the importance of further investigation into the role of the experimenter in ESP research.

Reference

1. Birge, W. R. A new method and an experiment in pure telepathy. *J. Parapsychol.*, 1948, **12**, 273-88.

Psychology Department
Duke University
Durham, North Carolina

LETTERS AND COMMENTS

PARAPSYCHOLOGY AND MARXIAN MATERIALISM[1]

Dear Dr. Rhine:

I have just finished reading *The Reach of the Mind* and I feel I must write and tell you the deep gratitude I feel toward you and all those who are working on the same subject.

While reading the book I couldn't help thinking back to a report of the late Zhdanov at a congress of Soviet philosophers and scientists which I happened to read at the beginning of last year. Zhdanov attacks violently the new orientation of Western (especially American) psychological and psychoanalytical studies. The reason is obvious: psychology and psychoanalysis (the more so parapsychology) want to bring man a better knowledge of himself and moral freedom. Instead of "decadent" Western psychoanalysis, Zhdanov recommends "autocriticism," which means a perpetual comparison of one's thoughts and actions with the standards of Marxist dogma in all fields—science, art, politics.

While you and your associates are working with a deep and sincere wish to get at the truth, to seize ultimate realities, there are countries in which the blessing of research for research's and truth's sake does not exist; where whatever one does, the dreadful question of his work's conforming to "scientific materialism" will always be put to him; where the value of one's work is measured, not by the distance separating it from truth, but by the distance separating it from conformity with a theory that has long since been proved false.

This scientific materialism, born in 1848, is by far the most important offspring of the materialistic and physicalistic current of thought of the past centuries. It has become a doctrine as intolerant as those of the most intolerant sects known in history. It has been propagated by different kinds of organizations in all countries, and after 1920 by a huge country and a powerful army.

The whole political, economic, and military organization of the

[1] This letter, the author of which must remain anonymous, was received from a member of the French Foreign Legion stationed in the Far East. He is an exile from one of the Balkan countries.—Ed.

Soviet empire is based on "scientific materialism" whose essential dogma is the absence of any nonmaterial principles in man and the universe, and its corollary, the cerebrocentric conception of man. Without this conception, historical materialism, collectivist economy, and political enslavement (the three components of Marxism) are nonsense.

Up to the last decade, materialism's great foe was Christian thought. The churches, though, failed to face the new need for logical and experimental evidence; so materialism won easy victories. But with the efforts of some persevering men to re-establish the psychocentric view of man by scientific methods, there appeared unexpectedly a new enemy far more dangerous to materialism, for these men used the very weapons of scientific investigation that had hitherto seemed the exclusive tool of materialism and physicalism.

I am sure that the progress of parapsychology is followed with great interest in Soviet scientific circles. Those responsible for the dissemination of Marxian doctrine must feel that the general and evident success of parapsychology means the future death (sooner or later, depending upon strategic data) of Marxism, which is synonymous with scientific materialism.

In the United States, England, and France anyone who wishes may buy a book on parapsychology, draw his own conclusions, and even make contributions to research. In countries dominated by Marxism, however, the student will never be tempted by seeing such books in the usual library. The impermeable wall against western ideas will not only be maintained, but will be reinforced, for new discoveries might jeopardize Marxian dogmas. The hope, then, of using the discoveries of parapsychology to forward human cooperation *stops* at a very precise limit: the boundaries of the Marxist-dominated countries. Since these countries occupy an appreciable part of the world, the unexpected conclusion is that the paths on which parapsychology may lead us are paths of war and not of peace. (I do not necessarily imply material war.) This conclusion may be shocking, but it is none the less inevitable, if two things are kept in mind: (1) the savage efforts which the leaders of the Soviet government have always made, and are ready to make, to keep themselves at the head of the Soviet empire, and (2) the fear (covered by pretentious Marxian argumentation) that the western ideal of

freedom of thought, if it should penetrate their fortress, could be for them the yeast of fatal developments. In fact they fear not only for their fortress, but also for all political, scientific and artistic strongholds they may have conquered all over the world. Think what a crushing defeat it will be for them when slowly but surely the masses of honest people, cheated by scientific materialism in any of its forms, will turn upon them. It will be impossible for the communistic staffs all over the world to keep the scientific truth about man's real nature from being known and finally accepted.

And so, though it may seem far from the arena of world-shaking events, parapsychology is really the master knot in the seemingly inextricable entanglement of theories, schools, dogmas, and disciplines which, divided into two great currents not always easy to determine, struggle to get and keep a leading position in the affairs of men.

Millions of people, like myself, driven from home in recent years or—in the impossibility of leaving—enduring the worst sort of vexations in the name of scientific materialism, will look most hopefully to any new evidence which may rise successfully against it. And indeed for the first time since scientific materialism was born, such evidence seems to be available. The latest conclusions of parapsychological research are in direct and irrefutable opposition to materialistic dogma. Perhaps it is hard to realize, even for you, what such an achievement means to the few of us victims of Marxism who know about it, and could mean to all others could they learn about it. You must not think that it is the simple and egocentric satisfaction of seeing a new difficulty arise for those who tortured, exiled, and ruined everyone who would not believe in their theories. No, it is the deep moral comfort of seeing that the truth (though undemonstrated), in which one grew up believing—that mind transcends the physical—is on the way to being scientifically established.

The hopes one had had about the existence of inner freedom, about survival, about the necessity of moral struggle—all had been covered with the sarcasm of subtle and apparently irrefutable scientific argumentation. The proponents of this argumentation had all the victories, be they in science or in politics. And so great was their success that those persons who still defended their hopes and beliefs had the discouraging impression that they were the last of a sinking world, and that the future belonged to materialism with

all its moral horror. Now this impression can be scientifically and logically dissipated. Maybe the conclusions to which parapsychology will lead us will not quite correspond to our former hopes and beliefs. It does not matter; they will certainly be on an evolutionary line with them. But one can be sure that they will give a mortal blow to materialism.

Though to you there seems to be a dramatic competition between atomic and human powers, to this far-away observer, there is none. To me it is most significant that the study of parapsychology has developed in the same countries as, and concomitant with, atomic research. Suppose parapsychology had developed in, say, Esthonia. This would not have been impossible, no great industrial installations being needed. What would be left of it after three invasions and the last occupation by the Red Army, when professors and students who would not accept Marxian constriction had to choose between being transported to Siberia and fleeing in little boats over 600 miles of Baltic Sea?

Cultural achievements must be defended, because whatever your hopes, there will for a long time be an enemy who is ready to destroy them. Write and say what you will, your words will stop at his frontiers; and your intentions, no matter how promising and honest, will never be known by the hundreds of millions of men and women toiling under his rule, brought up and trained to face one last and greatest struggle. The study of hidden powers in man, then, must necessarily proceed along with the development of material weapons for defending the whole cultural complex which generated such a study. Otherwise the entire work would be shattered to pieces by a materially stronger enemy, and the practical uses of parapsychology, if such be found, would be put to use for his own personal and crushing ambitions.

Nuclear physics and parapsychology are the two complementary poles of a magnificent culture, developed to an extent never known before, and conquering slowly unknown and marvelous regions. It is good to know that this culture has the weapons with which it can protect itself. Without them, any striving toward the unknown regions of mind, survival, and divinity will be meaningless for the future of humanity.

Sincerely yours,
Anonymous.

GLOSSARY

In order to avoid constant redefining of commonly recurring terms in papers appearing in this JOURNAL, the following definitions are submitted for convenient reference.

A simple description of the basic experimental methods and of the statistical procedures to be followed in evaluating ESP (both card tests and drawings tests) and PK may be found in the *Handbook of Tests in Parapsychology* by Betty M. Humphrey (published by the Parapsychology Laboratory, Duke University, Durham, N. C.).

AGENT: In tests for telepathy, the person whose mental states are to be apprehended by the percipient. In GESP tests, the person who looks at the stimulus object.

AVERAGE SCORE: Average number of hits per run.

CALL v.: To attempt to identify by ESP a stimulus object or a mental state of an agent.

CALL n.: The response described above; also the resulting selection.

CHANCE: The complex of undefined causal factors irrelevant to the purpose at hand.

CHANCE EXPECTATION = MEAN CHANCE EXPECTATION: The most likely score if only chance obtains.

CHANCE AVERAGE: Mean chance expectation expressed as an expected score, generally in terms of average per run.

CHI-SQUARE: A sum of quantities each of which is a deviation squared divided by an expected value. Also a sum of the squares of CR's.

CLAIRVOYANCE: Extrasensory perception of objective events as distinguished from telepathic perception of the mental state of another person.

CR (CRITICAL RATIO): A measure to determine whether or not the observed deviation is significantly greater than the expected random fluctuation about the average. The CR is obtained by dividing the observed deviation by the standard deviation. (The probability of a given CR may be obtained by consulting tables of the probability integral, such as Pearson's.)

CR OF THE DIFFERENCE: The observed difference between the average scores of two samples of data divided by the standard deviation of the difference.

DECK: Twenty-five ESP cards, five of each suit.

DEVIATION: The amount an observed number of hits or an average score varies from the mean chance expectation or chance average. A deviation may be total (for a series of runs) or average (per run).

DIE THROW: The elementary unit of data in a dice-throwing test represented by the throwing and reading of a single die, no matter how many dice are thrown at a time.

EMPIRICAL CONTROL: An experiment which wholly or partially follows the main experiment with the exception that the conditions are designed to exclude the possibility of the hypothesis being tested.

ESP (EXTRASENSORY PERCEPTION): Response to an external event not presented to any known sense.

ESP CARDS: Cards, each bearing one of the following five symbols: star, circle, three parallel wavy lines (called "waves"), square, plus.

ESP SYMBOLS: See plate opposite page 1, this JOURNAL, Vol. 1, March, 1937.

EXPECTATION; see CHANCE.

EXTRACHANCE: Not due to chance alone.

FREE MATERIAL: The stimulus objects of experiments in which an unlimited or unspecified range of stimulus objects is employed (as contrasted with methods such as card-calling in which the subject knows that the stimulus object is one of a known range).

FREE RESPONSE METHOD: An ESP test method in which a relatively unlimited range of stimulus objects is used and in which the subject is instructed to express freely as his response any range of thoughts and ideas.

GESP (GENERAL EXTRASENSORY PERCEPTION): A technique designed to test the occurrence of extrasensory perception, permitting either telepathy or clairvoyance or both to operate.

HIGH-DICE TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling eight or more.

LOW-DICE TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling six or less.

MEAN CHANCE EXPECTATION; see CHANCE.

P (PROBABILITY): A mathematical estimate of the expected relative frequency of a given event if chance alone were operative.

PARAPSYCHOLOGY: A division of psychology dealing with those psychical effects which appear not to fall within the scope of what is at present recognized law.

PERCIPIENT: The person who makes the calls in an ESP test.

PK (PSYCHOKINESIS): The direct influence exerted on a physical system by a subject without any known intermediate physical energy or instrumentation.

PRECOGNITION: Cognition of a future event which could not be known through rational inference.

PREFERENTIAL MATCHING: A method of scoring free responses. A judge ranks the stimulus objects (usually in sets of four) with respect to their similarity to, or association with, each response; and/or he ranks the responses with respect to their similarity to, or association with, each stimulus object.

PSI: A general term to identify personal factors or processes in nature which transcend accepted laws. It approximates the popular use of the word "psychic" and the technical one, "parapsychical."

PSI PHENOMENA: Occurrences which result from the operation of psi. They include the phenomena of both ESP (including precognition) and PK.

QD (QUARTER DISTRIBUTION): The distribution of hits in the record page (or in a logical subdivision thereof, such as the set or the half-set) as found in the four equal quarters formed by dividing the selected unit horizontally and vertically.

RUN: A group of trials, usually the successive calling of a deck of 25 ESP cards or symbols. In PK tests, 24 single die throws regardless of the number of dice thrown at the same time.

SCORE: The number of hits made in one run.

TOTAL SCORE: Total of scores made in a given number of runs.

AVERAGE SCORE: Total score divided by number of runs.

SD (STANDARD DEVIATION): Usually the theoretical root mean square of the deviations. It is obtained from the formula $\sqrt{npq}$, in which n is the number of single trials, p the probability of success per trial, and q the probability of failure. (For ESP cards, SD $= 2\sqrt{\text{no. of runs.}}$) Sometimes the SD used is that estimated from the observed variability in the scores.

SD OF THE DIFFERENCE: For both ESP cards and PK tests using dice, the SD of the difference is equal to $\sigma_s \sqrt{1/R_1 + 1/R_2}$ where σ_s is the SD of a single run and R_1 and R_2 are the number of the runs in the respective samples compared. This gives the SD of the difference for run score averages.

SERIES: Several runs or experimental sessions that are grouped in accordance with a stated principle.

SESSION: A unit of an ESP or PK experiment comprising all the trials of one test occasion.

SET: A subdivision of the record page serving as a scoring unit for a consecutive group of trials, usually for the same target.

SEVENS TEST: A PK technique in which the aim of the subject is to try to influence a pair of dice to fall with the two upper faces totaling seven.

SIGNIFICANCE: A numerical result is significant when it equals or surpasses some criterion of degree of chance improbability. Common criteria are: a probability value of .01 or less, or a deviation in the expected direction such that the critical ratio is 2.33 or greater.

SINGLES TEST: A PK technique in which the aim of the subject is to try to influence dice to fall with a specified face up.

STIMULUS OBJECT: The ESP card or drawing or other object, some identifying characteristic of which is to be apprehended by the percipient.

TARGET: In ESP tests, the stimulus object; or in telepathy, the mental state of the agent. In PK tests, the faces of the die (or combination of faces) which the subject attempts to bring up in the act of throwing.

TARGET CARD: The card which the percipient is attempting to perceive (i.e., to identify or otherwise indicate a knowledge of).

SUBJECT: The person who is experimented upon. In ESP tests, most commonly the percipient (though also the agent in GESP and telepathy). In PK tests, any individual whose task it is to influence the objects thrown.

TARGET DECK: The deck of cards the order of which the subject is attempting to identify.

TARGET FACE: The face on the die which the subject tries to turn up as a consequence of direct mental action.

TELEPATHY: Extrasensory perception of the mental activities of another person. It does not include the clairvoyant perception of objective events.

TRIAL: In ESP tests, a single attempt to identify a stimulus object. In PK tests, a single throw of the dice or other objects thrown.